THE SEARCH FOR EXPLICIT RECIPROCITY LAWS

- Quadratic and quartic irrationalities in the later books of Euclid.
- Cyclotomic irrationalities in Gauss's Disquisitiones.
- Higher reciprocity laws.
- Analytic and algebraic proofs in class field theory.
- Reciprocity laws and automorphic forms.
- Past and future uses of the trace formula, especially for $GL(2)$
- Algebraic number theory and the trace formula.
- Analytic number theory and the trace formula.

Introductory remarks

When invited to deliver these lectures I was urged not to devote all eight of them to my most recent elucubrations, whatever they might be, and assured that I was perfectly free to rehearse familiar material. I believe that Professor Varadarajan was referring to material familiar to me but perhaps unfamiliar to many members of the audience. I shall come to such material, but, being a cantankerous or at least contrary old codger, I have chosen to take his words in a sense contrary to that intended and to begin with material familiar to many of you, although not to all and certainly not to me.

The reciprocity laws appearing in my title occur on two levels, not necessarily of difficulty but of complexity. They all refer, not always so explicitly as one might like, to irrational numbers but these can be definite rational numbers, thus roots of single equations with rational coefficients, typical examples being quadratic or higher surds, in particular roots of unity. The more complex laws refer to collections of possibly irrational numbers such as the coordinates of the division points on an elliptic curve, or dually the coefficients of the equations defining covering curves; thus implicitly or usually explicitly, through the etale cohomology, they involve simultaneously both the geometric (expressed in the topology) and the arithmetic properties of the algebraic equations to be solved. Perhaps too fine a distinction should not be drawn, but I shall be concerned on the whole in these lectures with reciprocity laws at the first level.

What I would like to do is to take the occasion of this lecture to review the earlier stages of the development of these laws, thus of our understanding of the irrational, so that when we come to the last stages, those that pose still unsolved questions, questions that my generation is likely to bequeath as unfinished business to future generations, we will be able to formulate with at least minimal clarity what remains to be done.

The earlier stages seem to be well enough defined.

1) Quadratic irrationals: the Greeks.

2) Cubic and quartic irrationals: the Renaissance.

3) Beginnings of the general theory: Vandermonde and Lagrange and then Gauss's construction of periods in cyclotomic fields.

4) Kummer's theory of cyclotomic fields and of Kummer extensions.

5) Galois theory.

Of course, in the initial stages the theories of Galois and Kummer, who were born within a year or two of each other, were developed independently and with

quite different styles, penetrating philosophic insight on the one hand

Euclid and the irrational

Of course, Euclid and the irrational is a subject less broad than, say, the ancient Greeks and the irrational, but it is easier, as there is only one pertinent source, the thirteen books of Euclid's Elements of which, by and large, only two the 10th and the 13th have to do with the irrational. The 10th is perhaps deeper logically or philosophically but the 13th is more appealing geometrically and more immediately mathematical, but it is best to begin spending some time with the 10th, for here it is clear not so much with what kinds of irrationals the Greeks were familiar but rather with what kinds they were prepared to deal and for what kinds they had the most highly developed classification.

A good deal of material from the earlier books is of course implicit in the 10th book, especially from the theory of proportions, apparently due to Eudoxus, to which the 5th book is devoted and, in one way or another, the application of areas. This concept, appearing in Proposition 44 of Book 1, is the tool that allows the notion of ratio to be applied to two areas. It is a construction.

To a given straight line to apply, in a given rectilineal angle, a parallelogram equal to a given triangle.

Heath comments, "This proposition will always remain one of the most impressive in all geometrywhen account is taken (1) of the great importance of the result obtained ... and (2) of the simplicity of the means employed ... " He also observes that this is the proposition that allows, to use a concept familiar to us, the solution of quadratic equations.

The logical structure of Euclid is complex and fascinating. It is not my intention to discuss it here, but there are several propositions from Book VI, which treat of similar figures, especially triangles, and of the application of various areas, of which it is good to be aware when examining Book 10 and Book 13. The first, Proposition 19 assures us that, at first just for similar triangles, that the ratio of their areas is the square of the ratio of their sides. Formulated as in Heath's translation, it states,

Similar triangles are to one another in the duplicate ratio of the corresponding sides.

The propositions that, as explained by Heath, become when expressed algebraically the solutions of the quadratic equations

$$ax \mp \frac{b}{c}x^2 = S$$

are Propositions 28 and 29 of Book 6. The minus sign is Proposition 28, in which

it is understood that

$$S \leq \frac{c}{b}\frac{a^2}{4};$$

the plus sign is Proposition 29. All numbers are positive. In so far as numbers have to be constructible as lengths, it is as a result of these two propositions that the quadratic irrational numbers of Book 10 appear as very general entities in Euclid's mathematics. Specific numbers whose construction involves one of three specific irrational surds, namely $\sqrt{2}$, $\sqrt{3}$ and $\sqrt{5}$, appear, as is well known and as we shall recall, in other ways.

Taking, in particular, $a = 0$, $b = c = 1$, we construct with the help of Proposition 29, the square root of S. It is, however, not allowed in these two propositions to take $a = 0$. Simple surds appear as a result of Proposition 25, to wit,

To construct one and the same figure similar to a given rectilineal figure and equal to another given rectilinear figure.

Since Book 10 deals with ratios, a unit length δ is chosen. Thus any number S determines an area, that of the rectangle with sides δ and S. According to the proposition, there is a square of the same area. The side of this square has length (measured in terms of δ) equal to $\sqrt{S}$.

Heath begins his comments on this proposition by observing, "This is the highly important problem which Pythagoras is credited with having solved." He continues, referring to a passage from Plutarch (50 -125 A.D.) quoted earlier , "Among the most geometrical theorems, or rather problems, is the following: given two figures, to apply a third equal and similar to the other, on the strength of which discovery they say moreover that Pythagoras sacrificed. This is indeed unquestionably more subtle and scientific than the theorem which demonstrated that the square on the hypotenuse is equal to the squares on the sides about the right angle."

We can certainly ignore any tradition about sacrifices; we can also take any reference to Pythagoras with a grain of salt. I draw your attention, however, to the comparison between the two propositions and Plutarch's opinion about their relative merit and wonder whether you would express the same or the contrary opinion.

It is useful to state Propositions 28 and 29. It would be even more useful to explain how they are proved and how they are interpreted in Heath's sense. We do not have time for this, but you might ask yourselves how, and above all why, the Greeks arrived at these constructions.

Proposition 28. *To a given straight line to apply a parallelogram equal to a given rectilineal figure and deficient by a parallelogrammic figure similar to a given one: thus the given rectilineal figure must not be greater than the parallelogram described on the half of the straight line and similar to the defect.*

Proposition 29. *To a given straight line to apply a parallelogram equal to a given rectilineal figure and exceeding by a parallelogrammic figure similar to a given one.*

These two propositions are best explained with diagrams. For Proposition 28, the given rectilineal figure is C and it is its area that is pertinent. The figure D is the parallelogram to which the defect is to be similar. The line AB is the given line and ST is the applied parallelogram with deficit equal to the parallelogrammic QB, which is similar to D. For Proposition 28, the meanings of C and D does not change. AB is again the given line and the parallelogram AO is equal to C.

In spite of what I said and contrary to what most of us might expect, the surd $\sqrt{2}$ does not figure so prominently in Euclid as $\sqrt{3}$ and $\sqrt{5}$. The second of these two numbers appears in Proposition 30 of Book 6, thus

To cut a given finite straight line in extreme and mean ratio

Since numbers are necessarily ratios, take the line to be AB and its length to be 1. We cut it at C. Taking the length of AC to be x, we want

$$\frac{1}{x} = \frac{x}{1-x}$$

and $x > 1 - x$, for then the ratio of the whole line to the segment AC is equal to ratio of AC to CB. This is the meaning of cutting in extreme and mean ratio. The equation is $x^2 = 1 - x$ or $x^2 + x - 1$, with two solutions $x = 1/2 + \sqrt{5}/2$ and $y = 1 - x = 1/2 - \sqrt{5}/2$. Clearly $x > y$.

The proposition itself, in both its geometric form and its geometric interpretation, is an almost immediate consequence of Proposition 29. Another form of the same theorem is proved much earlier in Euclid, as Proposition 11 of Book 2.

To cut a given straight line so that the rectangle contained by the whole and one of the segments is equal to the square on the remaining segment.

The equivalence of the two propositions is, at least in their algebraic form, clear. Proposition 11 of Book 2 is used in Book 4, which deals largely with the construction of regular polygons, for the construction of a regular hexagon inscribed in a given circle. Since we know that $\sqrt{5}$ is contained in the cyclotomic field defined by the fifth roots of unity, we will hardly be surprised. Euclid does not construct explicitly a regular octagon, certainly because he felt it to be too easy, and the construction of a regular hexagon does not require explicitly the length of the sides of the six equilateral triangles into which it is naturally divided. So we see at this stage neither $\sqrt{2}$ nor $\sqrt{3}$ appearing. They appear later in the construction of the regular polyhedra inscribed in a sphere, the subject of Book 13.

Although Euclid's presentation sometimes appears at first glance haphazard, my experience suggests that this is never so. In Proposition 6 of Book 4, Euclid inscribes

a square in a circle but does not give the ratio of the side of the square to that of the circle. The ratio is of course $\sqrt{2}$. In Book 4 he does not explicitly inscribe an equilateral triangle in a circle; rather, in Proposition 2, he explicitly inscribes a triangle similar to an arbitrarily given triangle, which can, of course, be scalene. Only in Book 13 does he show that if the triangle is equilateral, it will have a side whose length is $\sqrt{3}$ times the radius of the circle.

Before turning to Book 13, in which specific irrationalities appear, it is best to examine, at least cursorily, Book 10, in which the sophistication of Euclid's understanding of quadratic and, to some extent of quartic irrationalities, which is, by inference if nothing else, also that of mathematicians contemporary with him. The material in the two books will presumably have been discovered in the course of the two centuries preceding him, but precise, or even approximate, dating discoveries or proofs is not my purpose here. Such dating, in so far as it is possible, is of course basic to any understanding at all of the rise of mathematics in ancient Greece.

The terms *rational* and *irrational*, used frequently in Book 10, have in Heath's translation a similar but different meaning than they have for us. The terms used in his translation for rational and irrational are *commensurable* and *incommensurable* and these are terms that apply to ratios. A significant proposition of Book 10 is Proposition 5.

Commensurable magnitudes have to one another the ratio which a number has to a number.

In other words the ratio defined by commensurable magnitudes is a common fraction.

A second significant proposition, much longer to state, is Proposition 10. Although proof given is unlikely to be that of the earlier Greek geometers, it establishes the existence of the first irrationals, the square roots of integers that are not squares. Thus, as part of the proposition we have the statement

. . . squares which have not to one another the ratio which a square number has to a square number will not have their sides commensurable in length . . .

The word *number* can be taken as referring to integers.

In Proposition 21 of Book 10, fourth roots are introduced, or rather square roots of square roots. The term employed in Heath's translation is *medial*. The proposition runs,

The rectangle contained by rational straight lines commensurable in square only is irrational, and the side of the square equal to it is irrational. Let the latter be called **medial.**

In modern language, this proposition asserts that the area of the rectangle is irrational. The language of Euclid is such that the second assertion of the proposition is equivalent to the first. As I observed the terms *rational* and *irrational* when they appear in Euclid have a different meaning than they have for us and refer as well to an assigned interval. A line is medial if the ratio of its length to the assigned length is the fourth root of a rational fraction that is not a square. So the notion of a quartic surd is clearly introduced in Euclid.

For reasons not clear to me, Euclid also introduces the notion of medial area. This is the area of a square on a medial line, thus the ratio of its area to that of the basic area, that of the square on the assigned line, is a quadratic surd. Then he formulates the proposition, Proposition 26 of Book 10, that the difference of two medial areas cannot be rational as follows,

A medial area does not exceed a medial area by a rational area.

I mention this because it is the first sign of an important principle of Kummer extensions, or if you prefer of Galois theory, namely that the fields $\mathbb{Q}(\sqrt{a})$ and $\mathbb{Q}(\sqrt{b})$ are linearly disjoint if a/b is not a square. What it asserts is weaker: that $\sqrt{a} - \sqrt{b}$ is not rational. It is understood that $a \neq b$. I do not find in Euclid a similar assertion for the sum. We shall return to Kummer extensions.

After medials, there are two classes of irrationals introduced and classified in Book 10, the *apotomes* and the *binomials*, although, I recall again, in Euclid it is not the numbers or ratios that are treated so much as the lengths they represent in terms of the fundamental length. These numbers are, respectively, differences or sums of a rational number and a quadratic surd or of two quadratic surds whose quotient is not rational.Thus they are numbers of the form $\rho \mp \sigma$ where each of ρ and σ is either rational or a quadratic surd or a quartic surd. There are two propositions along the lines of that just quoted: first of all, for binomials,

Proposition 42. *A binomial straight line is divided into its terms at one point only;*

then for apotomes,

Proposition 79. *To an apotome only one rational straight line can be annexed which is commensurable with the whole in square only.*

Thus, again in terms of numbers and not length, a given number can be expressed as $\rho \mp \sigma$, with σ and ρ as above, in only one way.

There are finer classifications of apotomes and binomials, and even further possibilities in which σ and ρ are allowed to be quartic surds. Book 10 is very long! There are also interpolations and extensions in which higher order surds are allowed and linear combinations with more than two terms, but it is time to pass to Book 13, but not without first presenting a judgement of De Morgan and a quite different

observation of my own.

De Morgan wrote, "Euclid investigates every possible variety of lines which can be represented by $\sqrt{(\sqrt{a} \pm \sqrt{b})}$, a and b representing two incommensurable lines ... This book has a completeness which none of the others (not even the fifth) can boast of; and we could almost suspect that Euclid, having arranged his materials in his own mind, and having completely elaborated the 10th Book, wrote the preceding books after it and did not live to revise them thoroughly."

Propositions 26, 42 and 79 can, of course, easily be established in modern terms and over any field not of characteristic 2. Suppose, for example, that a and b lie in F and that at least one of a and b is not a square. Let

$$c = \sqrt{a} + \sqrt{b} \neq 0.$$

Then c can not lie in F because there is at least one nontrivial automorphism σ of $F(\sqrt{a}, \sqrt{b})$ and

$$c = \sigma(c) = \pm\sqrt{a} + \pm\sqrt{b},$$

where not both signs can be $+1$. They can not both be -1 either for then $c = -c$ and $c = 0$. Suppose the first is $+1$ and the second -1. Then $\sqrt{a} = c$ and $b = 0$. This is out of the question.

Thus an identity of the form

$$5 = \sqrt{3 - 4i} + \sqrt{8 + 6i},$$

with $i^2 = -1$ is immediately suspect. The only possibility is that both $3 - 4i$ and $8 - 6i$ are squares, as indeed they are,

$$(2 - i)^2 = 3 - 4i, \qquad (3 + i)^2 = 8 + 6i.$$

I will observe later in connection with such identities that Renaissance mathematicians were, in a certain sense, less sophisticated than the ancients.

Book 13 is, in one respect, strongly related to Book 10 because the irrational figures in the construction of the regular polyhedra in a striking way. In another important respect, it is more closely related to Books 11 and 12 because all three deal with solid figures. Books 11 and 12 are, however, principally concerned with their volume, in particular, with the method of exhaustion.

As I have already observed, Book 13 does discuss, as preparation for answering similar questions about regular polyhedra, the lengths of the sides of polygons inscribed in a circle. Proposition 12, for example, states that

If an equilateral triangle be inscribed in a circle, the square on the side of the triangle is triple of the square on the radius of the circle.

There is no similar statement for a square because Euclid does not need it in his examination of regular solids.

The regular solids are systematically constructed, beginning with the *pyramid*, thus with the regular tetrahedron, in Proposition 13,

To construct a pyramid, to comprehend it in a given sphere, and to prove that the square on the diameter of the sphere is one and a half times the square on the side of the pyramid.

The statement is not entirely clear. A tetrahedron is constructed whose sides are equilateral triangles and which certainly is a regular tetrahedron if one exists. It is, on a moment's reflection, easy to see that it has all the desired symmetry, but this is not explicitly stated. What is explicitly proved or, at least evident from the construction, is that it has the required rotational symmetry around two of the four axes of the tetrahedron and that implies, of course, symmetry under the full tetrahedral group. The construction is as follows. In Figure 1, AB is a line whose length is the diameter of the given sphere. The point C is drawn so that AC is twice BC and ADB is a semicircle. The circle EFG has radius equal to BC and center H. A line HK is drawn perpendicular to the plane of this circle and of length AC. This defines a tetrahedron whose sides are those of the triangle EFG and the three lines joining E, F and G to K. By construction they all have the same length.

To show that all vertices of a sphere, we use Euclid's definition of a sphere as the figure obtained by rotating a semicircle about its diameter. The aesthetic flaw in this definition is that it defines a symmetric figure asymmetrically. It also reveals some of the weaknesses of the basic definitions of Euclid: of the straight line joining two points or of a circle. These do not concern us here. The definition's advantage is that it is convenient for the construction of regular figures. The diameter chosen is a line from K through H ending at L of length equal to that of AB. Thus HL and CB will be equal. As a consequence the semicircle with diameter KL in the plane KLE passes through E. On rotation this same semicircle will sweep out a sphere passing through G and F.

This is the construction. We have still to verify that the square on the diameter of the sphere is $1\frac{1}{2}$ times the square on the side of the tetrahedron. Certainly the length AB is triple the length CB or $3/2$ times the length AC. Since

$$\frac{AB}{AD} = \frac{DB}{DC} = \frac{AD}{AC},$$

the square of any of these numbers is $AB/AC = 3/2$. In particular, $AB^2 = 3AD^2/2$.

There is not time to give the constructions for the octahedron and the square, but I state the pertinent propositions leaving it to your curiosity either to discover the construction for yourself or to turn to Euclid.

Proposition 14. *To construct an octahedron and comprehend it in a sphere, as in the preceding case; and to prove that the square on the diameter of the sphere is*

double of the square on the side of the octahedron.

Proposition 15. *To construct a cube and comprehend it in a sphere, like the pyramid; and to prove that the square on the diameter of the sphere is double of the sphere on the side of the octahedron.*

The next two propositions, almost the final propositions in Euclid's thirteen books for there is only one more, show not the logical sophistication at which the Greeks had arrived only two centuries after Pythagorus but at least a part of the technical sophistication.

Proposition 16. *To construct an icosahedron and comprehend it in a sphere, like the aforesaid figures; and to prove that the side of the icosahedron is the irrational straight line called minor.*

If d is the diameter of the circle, then the square of the side of the pentagon is

$$\frac{d^2}{5}(10-2\sqrt{5}) = d^2(2-\frac{2}{\sqrt{5}}),$$

a number that may be written as

$$\left(d\sqrt{1+\frac{k}{\sqrt{1+k^2}}}-d\sqrt{1-\frac{k}{\sqrt{1+k^2}}}\right)^2 = d^2\left(2-2\sqrt{1-\frac{k^2}{1+k^2}}\right),$$

with $k=2$. So the side is

$$d\sqrt{1+\frac{k}{\sqrt{1+k^2}}}-d\sqrt{1-\frac{k}{\sqrt{1+k^2}}} = d\sqrt{2}\sqrt{1-\sqrt{1-\frac{k^2}{1+k^2}}}.$$

If d is rational in our sense, this is a quartic irrationality.

It is worthwhile to look again at the classification of irrationals in Book 10, for a *minor* is one of the irrational numbers met there. Here is the definition, expressed in Proposition 76 of Book 10,

If from a straight line there be subtracted a straight line which is incommensurable in square with the whole and which with the whole makes the squares on them added together rational, but the rectangle contained by them medial, the remainder is irrational; and let it be called **minor**.

This is easier for us to understand if we express it algebraically. Recall that the words rational and irrational do not necessarily have in Heath's translation the meaning that they have for us, or rather they do for areas but not for lengths. Thus if xr is the length of the initial straight line and yr the length to be subtracted, r

being the initially assigned length that turns all other lengths into numbers, then $x^2 + y^2$ is to be rational and y^2/x^2 irrational in our sense. The number xy is to be a quadratic surd. That is the meaning of medial for areas.

The existence of numbers with these properties is established in Proposition 33, which reads

To find two straight lines incommensurable in square which makes the sum of the squares on them rational but the rectangle contained by them medial.

Thus we are looking for x and y such that $x + y$ is rational, xy is a quadratic surd, and x^2/y^2 is irrational (a word always used in the contemporary sense in my text).

If $(xy)^2 = a$ and $x^2 + y^2 = b$, x^2 and y^2 are the roots of

$$z^2 - bz + a = 0,$$

where it is understood that a is not a square. Thus

$$z = \frac{b}{2} \pm \frac{\sqrt{b^2 - 4a}}{2} = \frac{b}{2}(1 \pm \sqrt{1 - 4a/b^2}),$$

and

$$x - y = \frac{\sqrt{b}}{\sqrt{2}}\sqrt{1 + \sqrt{1 - 4a/b^2}} - \frac{\sqrt{b}}{\sqrt{2}}\sqrt{1 - \sqrt{1 - 4a/b^2}}.$$

The number x^2/y^2 is equal to

$$\frac{1 + \sqrt{1 - 4a/b^2}}{1 - \sqrt{1 - 4a/b^2}} = \frac{(1 + \sqrt{1 - 4a/b^2})^2}{4a/b^2}.$$

The denominator is rational and the numerator is

$$2 - \frac{4a}{b^2} + 2\sqrt{1 - \frac{4a}{b^2}}.$$

So we need $1 - 4a/b^2$ not to be a square. Euclid takes it to be equal to

$$\frac{k^2}{1 + k^2},$$

where k is rational and $1+k^2$ is not the square of a rational number. More precisely, he takes $b = 1$, $a = 1/4(1 + k^2)$.

There seems to be no reason for this, except that he has established earlier that he can find a rational number k such that $1 + k^2$ is not a square. Of course, this choice, although not completely general, suffices for the expression of the side of a triangular face of the icosahedron.

I recapitulate part of the discussion leading to these numbers. Euclid proves two lemmas, of which the first is better known to you than the second, although the second is the one pertinent to the present discussion.

Lemma 1. *To find two square numbers such that their sum is also a square.*

Lemma 2. *To find two square numbers such that their sum is not a square.*

The first is proved with the usual pairs of integers $mnpq$ and $\frac{1}{2}(mnp^2 - mnq^2)$. The presence of so many factors, especially the factors m and n, is somewhat startling, but the construction of these Pythagorean triples had a history stretching back to the time of Pythagoras. So it is hardly surprising that it has become encrusted with what appears to us to be idle generality. I recall as well that Euclid's proof is not algebraic but geometric and is well worth examining on its own merits. Indeed his conception of the matter, especially of the notion of square, is also geometric and that accounts to some extent for the presence of m and n in the algebraic interpretation.

The second is proved by showing, as a consequence of the first, that

$$mnp^2mnq^2 + (\frac{mnp^2 - mnq^2}{2} - 1)^2$$

cannot be a square if

$$mnp^2mnq^2 + (\frac{mnp^2 - mnq^2}{2})^2.$$

is, the argument being essentially that the difference between this square and the square preceding it will have to be larger than the difference between these two numbers. The algebraic argument, given in Heath's comments, is easy to follow; the geometric argument more difficult, at least for me. It is the second lemma that leads to the particular construction of Proposition 33.

Euclid's construction of the icosahedron demands more comment than his construction of the dodecahedron. The dodecahedron was discovered earlier, perhaps because approximate forms of it appear naturally in iron pyrite crystals which would have, apparently, been known in the Iron Age. The diagrams that Heath includes in his commentary on Euclid's construction are much more transparent than Euclid's own figure, and make the role of the dual dodecahedron much more evident, at least of two faces of a dodecahedron similar to the dual dodecahedron. These are the faces $QRSTU$ and $EFGHK$ of his figure, dual to the vertices Z and X of the dodecahedron. Heath's figure understood, however, Euclid's becomes readily intelligible.

Anyhow, Euclid begins with a segment AB whose length is equal to the diameter of the sphere in which the icosahedron is to be inscribed. This he divides at C so that AC is four times BC. He constructs a circle $EFGHK$ with the radius BD.

The polygon $EFGHK$ he takes to be the regular pentagon constructed in Book 4. Then he bisects each of the arcs EF, FG and so on at the points L, M, N, O, P. He also translates the circle $EFGHK$ perpendicularly upwards a distance equal to the radius of the circle, obtaining another pentagon $QRSTU$

So he knows, from previous constructions, that QE, which is of length equal to BD, the radius of the circle, is the side of a regular hexagon inscribed in the circle. Now, he has already shown in Proposition 10 of Book 13, that the square on the side of the pentagon inscribed in a circle is equal to the sum of the squares on a decagon and a hexagon, thus – to verify it quickly in our terms –

$$|e^{2\pi i/10} - 1|^2 + |e^{2\pi i/6} - 1|^2 = |e^{2\pi i/5} - 1|^2,$$

which is equivalent to

$$3 - e^{2\pi i/10} - e^{-2\pi i/10} = 2 - e^{2\pi i/5} - e^{-2\pi i/5},$$

or, on rearrangement,

$$1 - \zeta + \zeta^2 - \zeta^3 + \zeta^4 = (1 + \zeta^5)/(1 + \zeta) = 0,$$

where $\zeta = e^{2\pi i/10}$. Euclid's demonstration is, of course, strictly geometrical with quite a different flavor.

Since the length of the side EP is that of a decagon inscribed in the given circle and the angle QEP is a right angle by construction, the length of QP is therefore that of a regular pentagon inscribed in the given circle. The same argument applies to UP. We conclude that QUP is an equilateral triangle. Continuing the argument, we see that the ten triangles along the central band in the figure between $EFGHK$ and $QRSTU$ are all equilateral.

Finally, Euclid constructs the line $XVWZ$ through the center V of the circle through $EFGHK$ and perpendicular to it, with XV and WZ equal to the side of a regular decagon in the circle and VZ to the side of a regular hexagon. The icosahedron so constructed will then, again thanks to of Proposition 10, have all its faces equilateral triangles.

Euclid then shows that it can be inscribed in the given sphere and shows that its side, or rather the ratio of its side to the radius of the sphere, is the ratio called minor. I find some confusion in Euclid between the definite article and the indefinite. In Book 10, the word minor refers to a type of irrational number, not to a specific irrational number of this type. In the statement of Proposition 16, a specific number is intended.

It is proved in Proposition 9 of Book 13 that ratio of the lengths of the sides of the regular hexagon and regular decagon is that of the two parts of a line cut in extreme and mean ratio. Thus

$$ZW : WV :: WV : ZV.$$

As a consequence,

$$ZV : VE :: EV : VX.$$

Since the angles ZVE and EVX are right, the angle XEZ will also be right. So the semicircle on XZ will pass through E and, for similar reasons, also through Q. Rotating it, we obtain a sphere passing through all of the vertices of the icosahedron.

Algebraically, Proposition 9 affirms that

$$\frac{|e^{2\pi i/6} - 1| + |e2\pi i/10 - 1|}{|e^{2\pi i/6} - 1|} = \frac{|e^{2\pi i/6} - 1|}{|e^{2\pi i/10} - 1|},$$

or

$$(1 + |e^{2\pi i/10} - 1|)|e^{2\pi i/10} - 1| = 1,$$

because $|e^{2\pi i/6} - 1| = 1$. In other words, $2\sin\frac{\pi}{10}$ is the positive quadratic irrationality satisfying $(1 + x)x = 1$ or $x^2 + x - 1 = 0$, so that $x = -1/2 + \sqrt{5}/2$. Since $2\sin\frac{\pi}{10} = 2\cos\frac{2\pi}{5}$ and $2\cos\frac{2\pi}{5} = \zeta + \zeta^4$ when $\zeta = e^{2\pi i/5}$. This follows from $1 + \zeta + \zeta^2 + \zeta^3 + \zeta^4 = 0$ and $(\zeta + \zeta^4)^2 = \zeta^2 + \zeta^3 + 2$. Once again, Euclid's proof is quite different.

It remains to show that the length XZ is equal to the length AB and to establish the length of the sides. Once again, this follows from earlier propositions in Book 13. By construction, BD and VW have the same length. So we have to establish that XZ has the same ratio to VW as AB has to BD. We agree with Euclid that it is enough to prove that their squares have the same ratio, in both cases 5. Since

$$AB : BD :: BD : CB,$$

we infer that

$$\frac{AB^2}{BD^2} = \frac{AB}{BC} = 5.$$

To establish that this is also the ratio of XZ to VW, Euclid has to use Proposition 3 of Book 13, which is pretty much a lemma included exactly for this purpose.

Proposition 3. *If a straight line be cut in extreme and mean ratio, the square on the lesser segment added to the half of the greater segment is five times the square on the half of the greater segment.*

Once again, I am going to verify this algebraically, just for orientation. The real challenge is rather to examine Euclid's proof and to understand what he and his predecessorsmight have been thinking as they constructed or verified the proofs available to them. Suppose, as before, that $x^2 = 1 - x$ and $x > 1 - x$. The claim of the proposition is that

$$(1 - x + \frac{x}{2})^2 = 5(\frac{x}{2})^2,$$

or

$$(1 - \frac{x}{2})^2 - \frac{5x^2}{4} = 0.$$

This is certainly clear upon simplifying.

We have already seen that by construction and by Proposition 9, the line VZ has been cut in extreme and mean ratio at W and ZW is its lesser segment. Thus if A' is the center of VW and thus the center of XZ, the square on ZA' is five times the square on $A'W$. since ZX is twice ZA' and VW twice $A'W$, the square on ZX is five times the square on VW.

The last lemma to which Euclid appeals is Proposition 11 of Book 13, which affirms,

If in a circle which has its diameter rational an equilateral pentagon be inscribed the side of the pentagon is the irrational straight line called minor.

In Proposition 11 as in Proposition 16, Euclid proves more than he affirms, the term *minor* being, as we have observed, only descriptive of a type of irrationality. What Euclid proves in Proposition 11 is that the side is $\sqrt{2 - 2/\sqrt{5}}$ times the diameter r or, expressed in a different way,

$$\frac{r}{2}\sqrt{5 + 2\sqrt{5}} - \frac{r}{2}\sqrt{5 - 2\sqrt{5}}.$$

It is his proof that would occupy us if we had more time, but in our concise algebraic language it means that

$$|e^{2\pi i/5} - 1|^2 = 2 - 2\cos 2\pi/5 = 5 - \sqrt{5}.$$

What he proves in Proposition 16 is that the edges of the icosahedron are obtained from the radius r of the circumscribed sphere by multiplying it by $\sqrt{2 - 2/\sqrt{5}}$. He states less than this. His argument runs as follows:

For, since the diameter of the sphere is rational, and the square on it is five times the square on the radius of the circle $EFGHK$, therefore the radius of the circle $EFGHK$ is also rational; hence its diameter is also rational.

Recall that the word *rational* is used in Heath's translation in connection with a given length to mean that the ratio of the length to the assigned length *or* of the square of the length to the square of the assigned length is the quotient of two integers. In any case, the radius of the circle $EFGHK$, from which the length of the edges will be deduced from Proposition 11 is the diameter d of the circumscribed sphere divided by $\sqrt{5}$ or the radius multiplied by $2/\sqrt{5}$. Euclid continues,

But (by Proposition 11), if an equilateral pentagon be inscribed in a circle which has its diameter rational, the side of the pentagon is the irrational straight line called minor.

And the side of the pentagon $EFGHK$ is the side of the icosahedron. Therefore the side of the icosahedron is the irrational square line called minor.

The exact length of the side in terms of the radius can be deduced immediately from Proposition 11.

According to Heath, there is at least one more construction of the icosahedron, due to Pappus in the late Hellenistic period (c. A. D. 300). In contrast to Euclid, whose construction is based on two parallel pentagons formed from edges of the icosahedron, Pappus bases his construction on four parallel circles, each containing three vertices of the icosahedron. Both constructions yield a great deal more information than the modern existence proof implicit, for example, in Hermann Weyl's book on symmetry and based solely on group-theoretical principles. There are also more recent explicit geometrical constructions.

The seventeenth and penultimate proposition in Book 13, the last book, is the construction of the regular dodecahedron.

To construct a dodecahedron and comprehend it in a sphere, like the aforesaid figures, and to prove that the side of the dodecahedron is the irrational straight line called apotome.

As Heath explains, the length of the edges is shown in effect to be $1/3(\sqrt{15} - \sqrt{3})$ times the radius of e sphere. The construction is based on a cube whose vertices are also vertices of the dodecahedron. I show Euclid's diagram as well as another offered by Heath that is more transparent. Unfortunately, we do not have enough time to examine this construction with any care.

The final proposition in Euclid is a comparison of the lengths of the sides of the five regular solids inscribed in a given sphere with a geometrical construction that gives them all. At the very end, as a remark that is not given formal status as a proposition, it is asserted that these are the only regular figures, indeed that these are the only polyhedra with faces given by a single regular polygon. Heath does not comment on it. The statement may be true, but the arguments are not convincing.

Gauss and cyclotomic irrationalities

We all know that Gauss was precocious and many of us are aware that he was especially fortunate not only in his teachers but also in the opportunities available to him in his youth: time and above all access to an excellent mathematical library. It is not certain what he read, but according to Buhler's biography, it can be assumed that he read some of the papers of Lagrange.

Gauss is, of course, a seminal figure in the transformation of the purely algebraic discoveries of the Renaissance period into the theory of equations, as created by, say, Galois and Abel, and the largely German algebraic number theory of the nineteenth century, but Lagrange and Vandermonde are critical transitional figures. So we will understand Gauss better if we have some familiarity with the papers of Lagrange and Vandermonde on the theory of equations, even if the understanding is only tentative and provisional because we cannot know exactly what he had read when, for example, he succeeded in finding the construction of the regular heptadecagon.

Stäckel in his appreciation of "Gauss als Geometer" that appears in vol.10″ of Gauss's *Werke* observes, however, that there is concrete evidence that he was familiar with Vandermonde's paper. On p. 58 of his article, he writes

Es ist sehr wahrscheinlich, dass GAUSS bei der Abfassung der Disquisitiones arithmeticae dessen (namely Vandermonde's) *Abhandlung gekannt hat, denn in dem Briefe an OLBERS vom 12. Oktober 1802 sagt er, dass wir über die Geometrie situs ≫nur einige Fragmente von EULER und einem von mir sehr hochgeschätzten Geometer VANDERMONDE haben≪.* Die Abhandlung über Geometria situs steht aber in demselben Bande der Pariser Denkschriften für das Jahr 1717 wie die Abhandlung über die Auflösung der algebraischen Gleichungen.

There is a reference to Stäckel's article in the chapter on the theory of equations in Bourbaki's *Éléments d'histoire des mathématiques*, a collection of essays written, I understand, largely by André Weil. This chapter contains an instructive review of the pertinent papers of Lagrange and Vandermonde.

Prof. Varadarajan has included in his monograph *Algebra in ancient and modern times* an excellent introduction to the theory of equations of third and fourth degree, a theory from which Lagrange starts. I shall pretty much take this theory for granted, although there is one point on which I would like to comment, as it pertinent to the theory of Kummer extensions to which we shall come in connection with class field theory.

Prof. Varadarajan gives the formula of Scipione del Ferro for the (single real) root of

$$X^3 + PX = Q, \qquad P, Q > 0,$$

namely

$$X = \sqrt[3]{\sqrt{\Delta} + \frac{Q}{2}} - \sqrt[3]{\sqrt{\Delta} - \frac{Q}{2}}, \qquad \Delta = \frac{Q^2}{4} + \frac{P^3}{27}.$$

He then applies it to the equation

$$X^3 + 6X = 20,$$

which obviously has the root 2, and concludes that

$$2 = \sqrt[3]{6\sqrt{3} + 10} - \sqrt[3]{6\sqrt{3} - 10},$$

because

$$\frac{Q}{2} = 10, \quad \Delta = 100 + 8 = 3 \times 36.$$

He refers to this as a remarkable identity.

Although in his text *New first course in the theory of equations* from which I learned about cubic equations, Dickson was careful to exclude equations with rational roots in order, as he explains at some length, not to confuse the students. So such factitious identities did not appear, I do remember having some presented to me. They made me uneasy, although I did not know why.

As we observed in connection with similar identities for quadratic irrationalities, Galois theory implies that the must be trivial. For if K is a field not of characteristic 3 that I suppose, since for our purposes there is no harm in adjoining them, contains the third roots of unity. Suppose $A \neq 0$, X and Y lie in K and

$$A = \sqrt[3]{X} + \sqrt[3]{Y}.$$

Then X and Y are necessarily cube roots in K. For, otherwise, KX, Y is a nontrivial Galois extension with a nontrivial automorphism σ and

$$A = \sigma A = \mu \sqrt[3]{X} + \nu \sqrt[3]{Y},$$
$$A = \sigma A = \mu^2 \sqrt[3]{X} + \nu^2 \sqrt[3]{Y},$$

with at least one of the two roots of unity μ and ν different from 1. Then neither can be 1 and they must be different. Since the Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & \mu & \nu \\ 1 & \mu^2 & \nu^2 \end{vmatrix} \neq 0,$$

we conclude that $A = X = Y = 0$, contrary to assumption.

Thus $6\sqrt{3} + 10$ must certainly be a cube root in the field obtained by adjoining the cube roots to the field $\mathbb{Q}(\sqrt{3})$ and, indeed, in this field itself. To see of what,

we calculate its norm which is $100 - 36 \times 3 = -8$. So if it is a cube, it is the cube of a number with norm -2, thus of a number ϖ that generates a prime ideal $\mathfrak{p}$ with $\mathfrak{p}\bar{\mathfrak{p}} = (2)$. There is an obvious such number, namely $1 + \sqrt{3}$ and its cube is, by good fortune,

$$1 + 3\sqrt{3} + 3 \cdot 3 + 3\sqrt{3} = 10 + 6\sqrt{3}$$

The cube root could have been harder to find, but never impossible.

Lagrange, in the long but discursively written paper *Sur la résolution algébrique des équations*, in which there is no sign of our crabbed modern style, analyzes the solution of Scipione del Ferro and Tartaglia and, in the course of his analysis, introduces some basic ideas of Galois theory. As Bourbaki (Weil?) observes, however,

Lagrange fait déjà la distinction entre les diverses fractions rationelles qu'on obtient à partir de V (a rational function of the roots of an equation) *par permutations des indéterminées* x_i *(*$1 \leq i \leq n$*), et les diverses valeurs qui prennent ces fractions lorsque les* x_i *sont les racines d'une équation algébrique à coefficients numériques donnés; mais il subsiste encore dans son exposé un certain flottement à ce sujet, et c'est seulement avec Galois que la distinction deviendra plus nette.*

The essential difference between Lagrange and Galois is perhaps that Lagrange, as we shall see, is inclined to work with all possible permutations of the roots and not just with those that preserve all relations between them over the field of the coefficients or over the ground field.

He begins with a speculation as to how Scipione del Ferro and Tartaglia may have arrived at their solution. Starting from

$$x^3 + mx^2 + nx + p = 0,$$

he observes as usual that m can be taken to be 0 and, trying what contemporary physicists would call an Ansatz, sets $x = y + z$, substitutes to obtain

$$y^3 + z^3 + p + (y + z)(3yz + n) = 0,$$

and thus demands that

$$y^3 + z^3 + p = 0, \qquad 3yz + n = 0.$$

In other words,

$$z = -\frac{n}{3y},$$

$$y^3 - \frac{n^3}{27y^3} + p = 0.$$

The second of these equations is equivalent to

$$y^6 + py^3 - \frac{n^3}{27} = 0.$$

This yields

$$y = \sqrt[3]{-\frac{p}{2} \pm \sqrt{\frac{p^2}{+}\frac{n^3}{27}}},$$

which when substituted in

$$x = y + z = y - \frac{n}{3y}$$

gives the formula of Scipione del Ferro.

Then Lagrange begins to reflect, first of all that there are six values for y but only three for x. This is explained by the invariance of $y - n/3y$ under $y \to -n/3y$ together with the relation

$$\sqrt[3]{-\frac{p}{2} + \sqrt{\frac{p^2}{4} + \frac{n^3}{27}}}\sqrt[3]{-\frac{p}{2} - \sqrt{\frac{p^2}{4} + \frac{n^3}{27}}} = -n/3.$$

Lagrange is more specific. He gives the six possible values of y as

$$\sqrt[3]{-\frac{p}{2} \pm \sqrt{\frac{p^2}{4} + \frac{n^3}{27}}}, \quad \alpha\sqrt[3]{-\frac{p}{2} \pm \sqrt{\frac{p^2}{4} + \frac{n^3}{27}}}, \quad \beta\sqrt[3]{-\frac{p}{2} \pm \sqrt{\frac{p^2}{4} + \frac{n^3}{27}}},$$

where α and $\beta = 1/\alpha$ are the two nontrivial roots of unity. If, for brevity, we set, now following the notation of Lagrange,

$$q = \frac{p^2}{4} + \frac{n^3}{27},$$

the three values of x are then

$$\begin{aligned}
&\sqrt[3]{-\frac{p}{2} \pm \sqrt{q}} + \sqrt[3]{-\frac{p}{2} \mp \sqrt{q}},\\
&\alpha\sqrt[3]{-\frac{p}{2} \pm \sqrt{q}} + \beta\sqrt[3]{-\frac{p}{2} \mp \sqrt{q}},\\
&\beta\sqrt[3]{-\frac{p}{2} \pm \sqrt{q}} + \alpha\sqrt[3]{-\frac{p}{2} \mp \sqrt{q}}.
\end{aligned}$$

It is understood that the signs are chosen coherently in each line.

He next takes a decisive step that leads him to the *Lagrange resolvent*. The paper begins on p.205 of vol. 3 of his *Oeuvres*. On p. 213 he writes

L'équation du sixième degré

$$y^6 + py^3 - \frac{n^3}{27} = 0$$

s'appelle la réduite du troisième degré, parce que c'est à sa résolution que se réduit celle de la proposée

$$x^3 + nx + p = 0.$$

Or nous avons déjà vu plus haut comment les racines de cette dernière équation dépendent des racines de celle-là; voyons réciproquement comment les racines de la réduite dépendent de celles de la proposée;

To do this he introduces the usual transformation, setting

$$x' = x - \frac{m}{3}$$
$$m' = n - \frac{m^2}{3}, \quad p' = p - \frac{mn}{3} + \frac{2m^3}{27}.$$

Thus

$$x'^3 + n'x' + p' = 0.$$

Then, repeating his previous explanations, he sets $x' = y' - n'/3y$ and takes r to be the cube root of

$$-\frac{p'}{2} + \sqrt{\frac{p'^2}{4} + \frac{n'^3}{27}},$$

so that three values of y are r, αr and βr. This gives as the three values of x',

$$x' = r - \frac{n'}{3r}, \quad x' = \alpha r - \frac{n'}{3\alpha r}, \quad x' = \beta r - \frac{n'}{3\beta r}.$$

Then, for brevity setting $s = n'/3r$, he gives the three values of x,

$$a = -\frac{m}{3} + r - s,$$
$$b = -\frac{m}{3} + \alpha r - \frac{s}{\alpha},$$
$$c = -\frac{m}{3} + \beta r - \frac{s}{\beta}.$$

These linear equations he does not leave to the reader to solve; he solves them for the reader. First of all,

$$a - b = (1 - \alpha)(r + \frac{s}{\alpha}),$$
$$a - c = (1 - \beta)(r + \frac{s}{\beta}),$$

or

$$\frac{\alpha(a - b)}{1 - \alpha} = \alpha r + s,$$
$$\frac{\beta(a - c)}{1 - \beta} = \beta r + s.$$

Thus, continuing to follow Lagrange,

$$r = \frac{\frac{\alpha(a-b)}{1-\alpha} - \frac{\beta(a-c)}{1-\beta}}{\alpha - \beta},$$

or expanding,

$$r = \frac{a}{(1-\alpha)(1-\beta)} + \frac{\alpha b}{(\alpha-1)(\alpha-\beta)} + \frac{\beta c}{(\beta-1)(\beta-\alpha)}.$$

Since the coefficients of this linear expression in a, b and c are all functions of the two nontrivial roots of unity, they can be simplified. Lagrange does not leave this to the reader either. From

$$X^3 - 1 = (X-1)(X-\alpha)(X-\beta)$$

he deduces on differentiation

$$3X^2 = (X-\alpha)(X-\beta) + (X-1)(X-\beta) + (X-1)(X-\alpha).$$

Setting successively $X = 1, \alpha, \beta$, he deduces

$$\begin{aligned} 3 &= (1-\alpha)(1-\beta), \\ 3\alpha^2 &= (\alpha-1)(\alpha-\beta), \\ 3\beta^2 &= (\beta-1)(\beta-\alpha), \end{aligned}$$

so that

$$r = \frac{a}{3} + \frac{b}{3\alpha} + \frac{c}{3\beta}.$$

Finally, appealing to the relation $\alpha\beta = 1$, he obtains

$$r = \frac{a + \beta b + \alpha c}{3}.$$

he prefers, for obvious notational reasons, since the equations did not turn out just as he wished to interchange α and β, which had not been precisely specified, and to write

$$r = \frac{a + \alpha b + \beta c}{3}.$$

Then he begins to explain the significance of this equation.

On voit d'abord par cette expression de y pourquoi la réduite est nécessairement du sixième degré; car comme cette réduite ne dépend immédiatement des racines a, b, c de la proposée, mais seulement des coefficients m, n, p, où les trois racines entrent

également, il est clair que dans l'expression de y on doit pouvoir échanger à volonté les quantités a, b, c entre elles; par conséquent la quantité y devra avoir autant de valeurs différents que l'on pourra former par toutes les permutations possibles dont les trois racines a,b, c sont susceptibles; or on sait par la théorie des combinaisons que le nombre des permutations, c'est-à-dire des arrangements différents de trois choses, est $3 \times 2 \times 1$*; donc la réduite en y doit être aussi du degré* $3 \times 2 \times 1$*, c'est-à-dire du sixième.*

There is more!

la même expression de y montre aussi pourquoi la réduite est résoluble à la manière des équations du second degré; car il est clair que cela vient de ce que cette équation ne renferme que les puissances y^3 *et* y^6*, c'est à dire des puissances dont les exposants sont multiples de* 3*; en sorte que, su r est une des valeurs de y, il faut que* αr *et* βr *en soient à cause de* $\alpha^3 = 1$ *et* $\beta^3 = 1$*; or c'est ce qui a lieu dans l'expression de y trouvée ci-dessus. Pour le faire voir plus aisément nous remarquerons que* $\beta = \alpha^2$*, car, puisqu'on a* $\alpha\beta = 1$ *et* $\alpha^2 - 1 = 0$*, on aura aussi* $\alpha\beta = \alpha^3$*, et de là* $\beta = \alpha^2$*; de sorte que l'expression de y pourra se mettre sous cette forme*

$$y = \frac{a + \alpha b + \alpha^2 c}{3},$$

d'où, en faisant toutes les permutations possibles des quantités a, b, c, on tire les six valeurs suivantes

(A)
$$\begin{gathered}\frac{a + \alpha b + \alpha^2 c}{3},\\ \frac{a + \alpha c + \alpha^2 b}{3},\\ \frac{b + \alpha a + \alpha^2 c}{3},\\ \frac{b + \alpha c + \alpha^2 a}{3},\\ \frac{c + \alpha b + \alpha^2 a}{3},\\ \frac{c + \alpha a + \alpha^2 b}{3}\end{gathered}$$

qui seront donc les six racines de la réduite. (Notice the length of the sentence) *Maintenant si l'on multiplie la première par* α*, et ensuite par* β *ou par* α^2*, on aura, à cause de* $\alpha^3 = 1$*, ces ceux-ci*

$$\frac{c + \alpha a + \alpha^2 b}{3} \quad et \quad \frac{b + \alpha c + \alpha^2 a}{3},$$

qui sont la sixième et la quatrième; et si l'on multiplie de même la seconde par α *et par* α^2*, on aura*

$$\frac{b+\alpha a+\alpha^2 c}{3} et \quad \frac{c+\alpha b+\alpha^2 a}{3},$$

qui sont la troisième et la cinquième. Il en sera de même si l'on multiple la troisième et la quatrième, ou la cinquième et la sixième par α *et par* α^2*, car on aura là également toutes les autres.*

At this point, we are at p. 217 of the memoir, which, I recall, began on p. 205. It continues with a calculation of the coefficients of the equation of sixth degree satisfied by the roots in (A). This is a matter of calculating various explicitly given symmetric expressions in a, b and c in terms of the elementary symmetric functions m, n and p. Then he examines the Tschirnhausian transformation and comments on papers of Euler and Bezout, before turning to equations of degree four and their *réduites*. In other words, as for equations of degree three , given an equation of degree four in x, he attempts, as we do, to reduce its solution to that of an equation of another unknown y, which has one or the other special properties. Basically he reviews the known possibilities. What seems to be new and different is that he expresses y in terms of the four roots x', x'', x''', x'''' of the original equation and explains how the degree of the equation for y depends on the symmetry of this expression. For example if $y = x'x'' + x'''x''''$, which is invariant under eight permutations, then the equation for y will have degree three, but if $y = x' + x'' - x''' - x''''$ which is invariant under four permutations and invariant up to sign under eight, then there will be six values for y that are "equal and of opposite sign" in pairs.

Having discussedequations of degree three and four in considerable detail and, so far as I can judge, from a new perspective, Lagrange turns to the discussion of the solution in radicals of equations of arbitrary degree. He considers from this perspective two techniques, that of Tschirnhaus, to which apparently those of Euler and Bezout are not dissimilar, and a different one, the use of Lagrange resolvents, to reduce the solution of such an equation to the extraction of radicals.

The Tschirnhausian transformation begins with the equation

$$x^\mu + mx^{\mu-1} + nx^{\mu-2} + \cdots = 0, \tag{B}$$

sets

$$x^{\mu-1} + fx^{\mu-2} + \cdots + y = 0,$$

thus takes y as a polynomial in x, so that y also satisfies an equation of degree μ, say

$$y^\mu + Ay^{\mu-1} + By^{\mu-2} + \cdots + V = 0,$$

in which, from elimination theory, all the coefficients will be symmetric functions of $f, g, \ldots$ of degree $1, 2, \ldots$. To make all but V equal to 0 will entail solving an equation of degree $1 \times 2 \times \cdots \times (\mu - 1)$, but will result in

$$y^\mu = -V,$$

so that we can then solve for y in radicals.

Lagrange's own method is of the same nature. On p. 332, thus well along in the paper, he sets

$$t = x' + \alpha x'' + \cdots + \alpha^{\mu-1} x^{(\mu)}, \tag{C}$$

where $x', x'', \ldots, x^{(\mu)}$ are the roots of (B). The equation whose roots are all values obtained from t by permuting the roots of (B) will have degree $1 \times 2 \times \cdots \times \mu$, but the equation for

$$\theta = (x' + \alpha x'' + \cdots + \alpha^{\mu-1} x^{(\mu)})^{\mu}$$

will have degree $1 \times 2 \times \cdots \times (\mu - 1)$ and $t^{\mu} = \theta$. So the result is the same as that from the Tschirnaus transformation.

It has the same problems of course, except that, in some cases at least, θ will be more explicit than V. Actually, he introduces

$$x' + yx'' + y^2 x''' + \cdots + y^{\mu-1} x^{(\mu)},$$

where y is any root of $y^{\mu} - 1 = 0$. These are now called Lagrange resolvents.

Lagrange's treatment of the effect of permutations on functions of the roots $x', x'', \ldots$ has much in common with Galois theory. In particular, he is aware that there may be relations between roots that are not preserved by all permutations and that this can have a decisive effect on whether and how it might be solved. Nevertheless, he does not apply his methods to equation in which such relations are present.

One of the most obvious is the equation for the primitive μth roots of unity, especially when μ is prime,

$$x^{\mu-1} + x^{\mu-2} + \cdots + x^{\mu} + 1 = 0. \tag{D}$$

Unless has some supplementary understanding of the equation, there is no reason to choose one of the possible Lagrange resolvents rather than another, and the resolvent does depend on the order in which the roots appear. We know that if $x = e^{2\pi i/\mu}$ is one root, then the others are x^k, $1 \leq k \leq \mu - 1$, but unless we pair them correctly with the $(\mu - 1)$st roots of unity $1, \alpha, \alpha^2, \ldots$ the $(\mu - 1)$st power of the resolvent (C) may not be especially simple. It will still, as Lagrange asserts, be the root of an equation of degree $\mu - 1$ whose coefficients can themselves be found by solving an equation of degree $1 \times 2 \times \cdots \times (\mu - 2)$, but we can do much better if we are careful.

We first examine the matter deploying all the concepts and facts with which we are familiar. If z is a root of (D) and $\gcd(k, \mu) = 1$, then z^k is also a root that, of course, only depends on k modulo μ. We know, moreover, that the Galois group is defined by the transformations $\sigma_k : \quad x \to x^k$, $1 \leq k \leq \mu - 1$. Moreover we can find an integer g such that the sequence $1, g, g^2, g^3, \ldots, g^{\mu-2}$ yields each nonzero residue

class modulo μ exactly once. Thus $x, x^g, x^{g^2}, \dots, x^{g^{\mu-2}}$ is the set of roots of (D). Suppose α is a primitive $(\mu-1)$st root of unity. We pair the roots of (D) and the $(\mu-1)$st roots of unity, thus the powers of α, so as to form

$$\varpi = \varpi_\alpha = x + \alpha x^g + \alpha^2 x^{g^2} + \cdots + \alpha^{\mu-2} x^{g^{\mu-2}}.$$

This number lies in a field larger than $\mathbb{Q}(x)$, but as $\mu-1$ and μ are relatively prime, we may extend σ_k to the larger field by letting it act trivially on α.

Then, for example,

$$\sigma_g : \quad \varpi \to x^g + \alpha x^{g^2} + \alpha^2 x^{g^3} + \cdots + \alpha^{\mu-2} x = \alpha^{-1}\varpi.$$

In the same way,

$$\sigma_{g^l} : \quad \varpi = \alpha^l \varpi.$$

Thus

$$\varpi^{\mu-1} = \prod_{l=0}^{\mu-1} \sigma_{g^L} \varpi \in \mathbb{Q}(\alpha)$$

There is no need to take a primitive $(\mu-1)$st root of unity. The number α may be replaced by $\beta = \alpha^f$ with $(\mu-1) = ef$. Then $\varpi_b eta^e$ lies in $\mathbb{Q}(\beta)$.

These ancillary roots of unity are, however, not necessary. We can free ourselves from them and thus from the Lagrange resolvant on observing that, for $ef = \mu - 1$, $h = g^e$, $\beta = \alpha^e$, the number ϖ_β is a linear combination of the *periods* introduced by Gauss,

$$(f, \lambda) = x^\lambda + x^{\lambda h} + x^{\lambda h^2} + \cdots + x^{\lambda h^{f-1}}, \qquad \lambda = 1, g, \dots g^{e-1}.$$

These numbers can be introduced for any λ that is not divisible by the prime μ, but there are only e different ones, $(f, \lambda) = (f, \lambda')$ if $\lambda' = h^l \lambda$. Gauss proves two theorems that I copy from the original. Together they show that the linear combinations over $\mathbb{Q}$ of the numbers $\{(f, \lambda)\}$, f fixed but λ arbitrary generate a field of degree e. They are not hard to prove. The art was to discover the construction and the significance.

The first states that the product of two periods with the same f is a sum of periods with the this f.

THEOREMA. *Sint* (f, λ), (f, μ) *duae periodi similes, identicae aut diversae, constetque* (f, λ) *e radicubus* $[\lambda]$, $[\lambda']$, $[\lambda'']$ *etc. Tunc productum ex* (f, λ) *in* (f, μ) *erit aggregatum* f *periodorum similium puta*

$$= (f, \lambda + \mu) + (f, \lambda' + \mu) + (f, \lambda'' + \mu) + etc = W$$

This is because

$$x^{\lambda h^{a+c}} x^{\mu h^c} = x^{(\lambda'+\mu)h^c}, \qquad \lambda' = \lambda h^a.$$

The possibility is admitted that one of the periods is $(f, 0)$, which is just the rational integer e.

The second states that every period with a given f is expressible as a polynomial of degree e in a given one with rational coefficients.

THEOREMA. *Supponendo, λ esse numerum per n non divisibilem, et scribendo brevitatis ergo p pro (f, λ) quaevis alia similis periodus (f, μ), ubi etiam μ per n non divisibilis supponitur, reduci poterit sub formam talem*

$$\alpha + \beta p + \gamma pp + \cdots + \theta p^{e-1}$$

ita ut coefficientes α, β etc sint quantitates determinatae rationales.

Taking the possible periods with a given f and $\lambda \not\equiv 0 \bmod \mu$ to be the given one p, together with $p', p'', \ldots$, Gauss observes first that

$$0 = 1 + p + p' + p'' + \ldots, \tag{E$'$}$$

and then that, as a result of the first theorem applied repeatedly, there are relations

$$\begin{aligned} 0 &= p^2 + A + ap + a'p' + a''p'' + a'''p''' + \ldots \\ 0 &= p^3 + B + bp + b'p' + b''p'' + b'''p''' + \ldots \\ 0 &= p^4 + A + bp + b'p' + b''p'' + b'''p''' + \ldots \end{aligned} \tag{E$''$}$$

If we go up to the $(e-1)$st power, this will give us $e-1$ equations, from which we can eliminate the $e-2$ linear variables $p'', p''', \ldots$ to obtain a relation

$$0 = \mathfrak{A} + \mathfrak{B}p + \mathfrak{C}p^2 + \cdots + \mathfrak{M}p^{e-1} + \mathfrak{N}p',$$

in which not all coefficients are 0. We need to show, however, that $\mathfrak{N} \neq 0$.

If it were, then p would satisfy an equation of degree $e-1$. There is, however, the symmetry given by, for example, $\lambda \to g\lambda$ that is carried over to the equations (E), so that all periods would satisfy this equation. Since there are e of them, two must be equal. The relation $p - p' = 0$, when written out in terms of $x = e^{2\pi i/\mu}$ is a polynomial of degree at most $\mu - 1$ that vanishes at 1. Dividing by the factor $x - 1$, we obtain a polynomial of degree at most $\mu - 2$ that x satisfies This contradicts the irreducibility of the cyclotomic equation, already proven by Gauss.

I now recall very briefly, as it was the first striking consequence of this analysis of the cyclotomic equations, how the periods can be used to construct a regular polygon of $\mu = 17$ sides. As we know, this is a matter of establishing that the number x can be obtained by repeated extraction of square roots. Since it generates

a normal field of degree 16 over $\mathbb{Q}$ with Galois group cyclic, this is clear, but it is more interesting to see how this works concretely and, indeed, although there will be no more geometric constructions in these lectures, to see the explicit construction.

We may take $g = 3$. Taking $e = 2$, $f = 8$ gives two periods $(8, 1)$ and $(8, 3)$ that are, according to our review of Gauss's arguments, conjugate quadratic irrationalities over $\mathbb{Q}$. In fact, the pertinent equation can easily be calculated and they are found to be $(-1 \pm \sqrt{17})/2$. Calculating these two expressions and $(8, 1)$ approximately, we find that

$$(8, 1) = \frac{-1 + \sqrt{17}}{2}.$$

Then we have $(4, 1)$, $(4, 3)$, $(4, 9)$, $(4, 10) = (4, 27)$. Calculating as in the proofs of the two theorems, we see that $(4, 1)$ satisfies the equation

$$(4, 1)^2 - (8, 1)(4, 1) - 1 = 0,$$

so that

$$(4, 1) = \frac{(8, 1) \pm \sqrt{(8, 1)^2 + 4}}{2}.$$

Since $(8, 1)^2 = -(8, 1) + 4$ the expression under the square root is 8-(8,1) We find now that

$$(4, 1) = \frac{(8, 1) + \sqrt{8 - (8, 1)}}{2}.$$

I omit the similar calculations for (2,1), which is

$$e^{\frac{2\pi i}{17}} + e^{-\frac{2\pi i}{17}} = 2\cos 2\pi/17,$$

because $3^8 \equiv 81^2 \equiv (-4)^2 \equiv -1$ modulo 17, and is enough to construct the regular heptadecagon. The result may be written in several ways. One is

$$\frac{-1}{8} + \frac{\sqrt{17}}{8} + \frac{\sqrt{34 - 2\sqrt{17}}}{8} + \frac{\sqrt{68 + 12\sqrt{17} - 16\sqrt{34 + 2\sqrt{17}} - 2(1 - \sqrt{17})\sqrt{34 - 2\sqrt{17}}}}{8}.$$

Gauss presents it in a more elegant form.

I have not had the time to reflect on the context in which Gauss was able to construct the regular heptadecagon, but there are clues in the report of Stäckel *Gauss als Geometer* published, along with many other appreciations in Gauss's *Werke*. In particular, as a footnote to the remark quoted at the beginning, he writes,

CH. A. VANDERMONDE, *Remarques sur les problèmes de situation*, Histoire de l'Acad., année 1771, Paris 1774, Mémoires, S. 566; (This will be the paper on analysis situs.) *Sur la résolution des équations*; ebenda, S. 365; die letztere Abhandlung ist in deutscher Sprache herausgegeben von C. ITZIGSOHN, VANDERMONDE,

Abhandlungen aus der reinen Mathematik, Berlin, 1887. Auf S. 375 behauptet VANDERMONDE, die Gleichung $x^n - 1 = 0$ sei für jeden Grad n durch Wurzelziehen lösbar und führt diem Rechnungen für einige Fälle durch, im Besonderen für $n = 11$. Für die Exponenten $n \leq 10$ hatte schon EULER, *De extractione radicum ex quantitatibus irrationabilis*, Comment. acad.sc. Petrop. 13 (1741/3), 1751, §39 bis 48, Opera omnia, ser. I, vol. 6, §31, die Wurzeln mittels blosser Wurzelziehungen dargestellt; dagegen meint er, führe der Fall $n = 11$ auf eine Gleichung fünften Grades, deren Lösung noch verborgen sei.

It is at first glance not clear what the difficulty is, as the solutions of $x^\mu - 1 = 0$ can certainly be obtained by the extraction of roots. They are roots, the μth roots of unity. So far as I can see on glancing at Euler's paper is that what is wanted is to express the real numbers, $\cos 2\pi/\mu$ and $\sin 2\pi/\mu$ in terms of real roots of real numbers. Euler gives, for example, a list of x for $\mu = 9$.

$$
\begin{aligned}
&= 1 \\
&= \frac{-1+\sqrt{-3}}{2} \\
&= \frac{-1-\sqrt{-3}}{2} \\
&= \sqrt[3]{\frac{-1+\sqrt{-3}}{2}} \\
&= \frac{-1+\sqrt{-3}}{2}\sqrt[3]{\frac{-1+\sqrt{-3}}{2}} \\
&= \frac{-1-\sqrt{-3}}{2}\sqrt[3]{\frac{-1+\sqrt{-3}}{2}} \\
&= \sqrt[3]{\frac{-1-\sqrt{-3}}{2}} \\
&= \frac{-1+\sqrt{-3}}{2}\sqrt[3]{\frac{-1-\sqrt{-3}}{2}} \\
&= \frac{-1-\sqrt{-3}}{2}\sqrt[3]{\frac{-1-\sqrt{-3}}{2}}
\end{aligned}
$$

Indeed, he gives lists up to $\mu = 10$, but not for $\mu = 11$, because this leads to an equation of fifth degree. Apparently the list is to be found in Vandermonde's paper, but I have not yet seen it. We, of course, know how to go about creating it for $\mu = 11$ with the help of Gauss's periods.

Class fields

What we have seen in Euclid and Gauss is the explicit construction, and to some extent, especially in Euclid, classification, of special kinds of irrational algebraic numbers. We could ask for a classification of all algebraic numbers. *Class field theory* provides this in a certain sense for all numbers that lie in *abelian* extensions of a given number field. Thus the classification is with respect to an assigned number field and the classification is less of the individual numbers in the extensions than of the fields. So we are certainly working at a much higher level of sophistication then Euclid, and, as you shall see, even than Gauss, for the classification requires not only notions of Galois theory but also the ideal numbers (or, in modern terminology, simply ideals) introduced by Kummer.

This and other notions we will have to take to some extent for granted, but it is instructive to glance at some of the early calculations, those of Kummer, who was one of the first to examine cyclotomic fields after Gauss, not the very first and, when he began, not the most careful but ultimately the most thorough. I would like to indicate some of the elements in the proofs, but I have no desire to present a systematic argument. Sometimes I draw on our extensive baggage of modern notions and modern results, but sometimes I shall adhere to Kummer's more direct approach that relies more on basic algebra. The purpose is to acquire a feel for some of the earliest fields of algebraic numbers to be investigated, to understand them as class fields, and at the same time to introduce ourselves to the ideas of Kummer himself.

It will be convenient to use the notation of the four early papers to which we occasionally eventually refer:

De numeris complexis, qui radicubus unitatis integris realibus constant,

Über die Divisoren gewisser Formen der Zahlen, welche aus der Theorie der Kreistheilung entstehen,

Zur theorie der complexen Zahlen,

Über die Zerlegung der aus Wurzeln der Einheit gebildeten complexen Zahlen in ihre Primfactoren.

The last is a formal presentation of a complete theory; the other three are more in the way of appealing introductions to Kummer's new ideas. The first, presented to the University at Königsberg on the occasion of its third centennial jubilee, is largely an account of his first calculations and there is much to recommend beginning with it.

Suppose, to use Kummer's notation, that λ is a prime and α is a primitive λth

root of unity. The irreducible equation over $\mathbb{Q}$ satisfied by α is

$$\Phi(x) = x^{\lambda-1} + x^{\lambda-2} + \cdots + x + 1 = 0. \tag{A}$$

We consider the ring $R = \mathbb{Z}(\alpha)$, which is, as implicitly understood by Kummer, the ring of integers in the field $K = \mathbb{Q}(\alpha)$.

If p is congruent to 1 modulo λ, then the left side of the equation (A) factors completely modulo p,

$$x^{\lambda-1} + x^{\lambda-2} + \cdots + x + 1 \equiv (x - u_1) \ldots (x - u_{\lambda-1}) \pmod{p},$$

where the u_i, $i = 1, \ldots, \lambda - 1$ are just the λth roots of unity modulo p. Thus, if $\mathfrak{p}$ is any prime divisor of p in K,

$$\prod_{i=1}^{\lambda-1} (\alpha - u_i) \equiv 0 \pmod{\mathfrak{p}},$$

because

$$\prod_{i=1}^{\lambda-1} (\alpha - u_i) \equiv 0 \pmod{p}.$$

Thus for some unique $u = u_i$, $\alpha - u \in \mathfrak{p}$. In particular, $\mathfrak{p}$ is of degree 1 over p. In other words, $\mathbb{Z}_p = \mathbb{Z}/(p)$ which is contained in $R/\mathfrak{p}$ is actually equal to it and the norm $N\mathfrak{p}$ is equal to p. Moreover $\mathfrak{p} = (p, \alpha - \eta)$.

To each of the conjugates $\alpha_1, \ldots, \alpha_{\lambda-1}$ of α – among which we include α itself – there is associated in the same way one of the u_i. Different α_i must be associated to different u_i, because $u_i - u_j$, $i \neq j$ is prime to p. Indeed, setting $\alpha_0 = 1$, we have

$$\prod_{\substack{0 \leq i,j \leq \lambda \\ i \neq j}} (\alpha_i - \alpha_j) = \pm \begin{vmatrix} 1 & 1 & \ldots & 1 \\ \alpha_0 & \alpha_1 & \ldots & \alpha_{\lambda-1} \\ \alpha_0^2 & \alpha_1^2 & & \alpha_{\lambda-1}^2 \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \alpha_0^{\lambda-1} & \alpha_1^{\lambda-1} & & \alpha_{\lambda-1}^{\lambda-1} \end{vmatrix}^2 = \pm \lambda^\lambda \tag{B}$$

Thus we can number the η_j so that $\mathfrak{p} = (p, \alpha_j - \eta_j)$ for all j. Consequently, $\mathfrak{p}_j = (p, \alpha - \eta_j) = \sigma\mathfrak{p}$ if $\sigma\alpha_j = \alpha$ are $\lambda - 1$ different, indeed relatively prime to each other, ideals dividing p with

$$(p) = \prod_{j=1}^{\lambda-1} \mathfrak{p}_j \tag{C}$$

and with $p = \mathrm{N}\,\mathfrak{p}$ or $(p) = \mathrm{N}\,\mathfrak{p}$, according to the way the norm is viewed.

There are various ways to persuade oneself of (C), according as to how much general theory one assumes. In particular, we can observe that the polynomials

$$P_j(x) = \prod_{k \neq j} (x - u_j)$$

over the finite field with p elements have no common root. We can therefore find polynomials $Q_j(x)$ such that

$$\sum_j Q_j(x) P_j(x) \equiv 1 \pmod{p}.$$

Then, for any integer c,

$$\prod_j (\alpha - u_j + pcQ_j(\alpha))$$

is certainly in $\prod \mathfrak{p}_j$ and is congruent to

$$\Phi(\alpha) + bp + cp \sum Q_j(\alpha) P_j(\alpha) \equiv (b + c)p \pmod{p^2}.$$

because $\Phi(\alpha) = 0$. We can certainly arrange that this expression is not divisible by p^2. On the other hand, the ideal on the right of (B) clearly contains $p^{\lambda - 1}$.

If the field $\mathbb{Q}(\alpha)$, or rather the ring $\mathbb{Z}(\alpha)$, were a unique factorization domain, then the ideal $\mathfrak{p}$ would be principal $\mathfrak{p} = (\varpi)$ and $p = \pm \mathrm{N}\, \varpi$. On this assumption, Kummer gives in the Konigsberg paper, for $5 \leq \lambda \leq 23$, a ϖ for each $p \equiv 1 \pmod{\lambda}$ and not greater than 1000, except, of course, when no such ϖ exists. This occurs only for $\lambda = 23$ because that is the first prime for which $\mathbb{Z}(\alpha)$ is not a principal ideal domain. The p for which ϖ cannot be found are 47, 139, 277, 461 and 967.

Since $\mathbb{Q}(\alpha)$ is, in a number or respects, one of the simplest examples of a class field, it is useful to explain how primes that are not congruent to 1 modulo λ factor. Suppose f is the smallest positive power of p for which $p^f \equiv 1 \pmod{\lambda}$. Then f divides $\lambda - 1$, $\lambda - 1 = ef$. If $p \equiv g^r \pmod{\lambda}$ then $e | r$. We have just examined the case, $f = 1$, $e = \lambda - 1$.

In general, Kummer considers the e periods introduced by Gauss,

$$(f, \mu) = \alpha^{\mu} + \alpha^{\mu h} + \cdots + \alpha^{\mu h^{f-1}}.$$

For convenience, I have replaced the notation of Gauss by that of Kummer. So x has become α; Gauss's λ is now μ; and Gauss's n is now λ. Moreover $h = g^e$. I recall that g is a generator of the group of nonzero residues modulo λ. Following Kummer but with a small modification, we now denote the periods by $\eta_1, \eta_2, \ldots, \eta_e$ with $\eta_i = (f, g^i)$, $i = 0, \ldots, e - 1$. We already saw, following Gauss, that these periods satisfy an equation of degree e with rational coefficients. The equation may be obtained from

$$\varphi(y) = (y - \eta_1)(y - \eta_2) \ldots (y - \eta_e) = 0$$

on expanding out and using Gauss's theorem on products of periods. As a result the coefficients of φ will each be an integer plus an integral linear combinations of periods. Since they are rational they are each integral and the coefficient of highest order is, of course, 1.

We have modified the notation of Kummer slightly in order to conform to our treatment for $f = 1$. We now let η be any one of the η_i. There is a difference between $e = 1$ and $e > 1$. For $e = 1$, when the η_i are periods of length 1 and each equal to a root α_i of unity, we saw that

$$\begin{aligned}\prod_{\substack{0\le i,j\le\lambda\\ i\ne j}}(\alpha_i-\alpha_j) &= \pm\prod_i(1-\alpha_i)^2\prod_{\substack{1\le i,j\le\lambda\\ i\ne j}}(\alpha_i-\alpha_j)\\ &= \pm\lambda^2\prod_{\substack{1\le i,j\le\lambda\\ i\ne j}}(\alpha_i-\alpha_j)\end{aligned}$$

was a power of λ and thus not divisible by the primes of interest to us, namely primes different from λ. So the same is true of

$$\prod_{\substack{1\le i,j\le\lambda\\ i\ne j}}(\alpha_i-\alpha_j) = \prod_{\substack{1\le i,j\le\lambda\\ i\ne j}}(\eta_i-\eta_j) \tag{D}$$

For a general e this may not be so. The factor may be divisible by a finite number of primes in addition to λ. For the moment I exclude them.

The expression (D) is of course the discrimant of the monic polynomial $\varphi(y)$ and is, as we all remember, expressible as a universal polynomial with integral coefficients in the elementary symmetric functions of the coefficients of φ. If p is not one of the excluded primes, as we now assume, then it does not divide the discriminant. Since the polynomial is universal, the discriminant of $\varphi \mod p$ is the residue of the discriminant of φ and not 0. So $\varphi \pmod p$ has distinct roots.

Now we can argue as before. There are two fields in addition to $\mathbb{Q}$ to be considered, the field $K = \mathbb{Q}(\alpha)$ and the field

$$k = \mathbb{Q}(\eta) = \mathbb{Q}\eta_1 + \cdots + \mathbb{Q}\eta_e.$$

Suppose $\mathfrak{P}$ is a prime divisor of p in K and $\mathfrak{p}$ the intersection of $\mathfrak{P}$ with k.

Suppose $p \equiv g^r \pmod\lambda$. Since

$$\eta_i^p \equiv \alpha^{pg^{i-1}} + \alpha^{pg^{e+i-1}} + \cdots + \alpha^{pg^{(f-1)e+i-1}} \pmod{\mathfrak{p}},$$

we conclude that $\eta_i^p \equiv \eta_{r+i} \equiv \eta_i \pmod{\mathfrak{p}}$ when $e|r$. In other words each η_i is a root of $y^p - y$ modulo $\mathfrak{p}$. Thus there is an integer u_i such that $\eta_i - u_i \equiv 0 \pmod{\mathfrak{p}}$ and $u_i - u_j$ is not congruent to 0 modulo $\mathfrak{p}$ if $i \ne j$ because $\eta_i - \eta_j$ is not. This was the assumption that p did not divide the discriminant.

If η is η_i and u is u_i, then $\mathfrak{p} = (\eta - u, p)$ and it is a prime ideal of degree one in k because every algebraic integer in k, namely every element in $\mathbb{Z}\eta_1 + \cdots + \mathbb{Z}\eta_e$ is congruent to an ordinary integer modulo $\mathfrak{p}$. The conjugates of $\mathfrak{p}$ are the ideals $(\eta_i - u, p)$ and we can argue as before that

$$(p) = \mathrm{N}\,\mathfrak{p} = \prod_{i=1}^{e} \mathfrak{p}_i.$$

The ideal $\mathfrak{p}$, and thus each of its conjugates, remains prime in K. As usual I use the same notation for the ideal $\mathfrak{p}$ of k and the ideal $R\mathfrak{p}$ of K, where, as before, $R = \mathbb{Z}(\alpha)$ is the ring of integers in K. Observe first that $R \cap k = \sum_i \mathbb{Z}\eta_i$ modulo $\mathfrak{p}$ is just $\mathbb{Z}_p$ because each η_i is congruent to the integer u_i modulo $\mathfrak{p}$. Thus

$$R \cap k \pmod{p} = \oplus_i R \cap k \pmod{\mathfrak{p}_i}$$

is a direct sum of copies of $\mathbb{Z}_p$. Then

$$R \pmod{p} = \oplus_i R \pmod{\mathfrak{p}_i},$$

where here $\mathfrak{p}_i$ is to be taken as an ideal in K. Because the ideals $\mathfrak{p}_i$ are conjugate, the quotients on the right all have the same dimension over $\mathbb{Z}_p$. Since there are e of them, this must be $f = (\lambda - 1)/e$. Each of the quotients is a direct sum of fields, because the left side is, for the discriminant of

$$x^{\lambda-1} + x^{\lambda-2} + \cdots + x + 1 = 0$$

is a power of λ and therefore not divisible by p. Since p^f is the smallest power of p such that λ divides $p^f - 1$, the finite field of order p^f is the smallest field of characteristic p containing the λth roots of unity. Since each of the summands into which $R \pmod{\mathfrak{p}_i}$ decomposes contains these roots, each is of degree at least f over $\mathbb{Z}_p$. Thus each summand is a field and $\mathfrak{p}_i$ is prime. In particular, $\mathfrak{p}$ is prime.

Since Kummer was born in 1810 and Galois in 1811, these arguments, which employ such a liberal use of finite fields, are unlikely to have been used by Kummer. Nor will Kummer have used our notion of ideal. His concept of *ideal number*, which the arithmetic of cyclotomic fields forced him to introduce, yielded of course an equivalent theory as ours and not less sophisticated, but its sophistication was of a different nature. It is instructive to examine his comments and his definitions.

Recall that he was examining numbers of the form

$$f(\alpha) = a + a_1\alpha + a_2\alpha^2 + \cdots + a_{\lambda-1}\alpha^{\lambda-1}.$$

His first impulse will have been to repeat the definition with which we are familiar from the usual theory of prime numbers, namely that a number is prime if it does not factor.

However he writes, in the third of the papers listed,

Eine solche comlexe Zahl kann entweder in Factoren derselben Art zerlegt werden; oder auch nicht. Im ersten Fall ist sie eine zusammengesetzte Zahl; im andern Fall ist sie bisher eine complexe Primzahl genannt worden. Ich habe num aber bemerkt, daß wenn auch $f(\alpha)$ auf keine Weise in complexe Factoren zerlegt werden kann, sie deshalb noch nicht die wahre Natur einer complexen Primzahl hat, weil sie schon gewöhnlich der ersten und wichtigsten Eigenschaft der Primzahlen ermangelt: nämlich, daß das Product zweier Primzahlen durch keine von ihnen verschiedene Primzahl theilbar ist. Es haben vielmehr solche Zahlen $f(\alpha)$, wenn gleich sie nicht in complexe Factoren zerlegbar sind, dennoch die Natur der zusammengesetzten Zahlen; die Factorer sind aber alsdann nicht wirkliche, sondern ***ideale complexe Zahlen****.*

He goes on to comment that the introduction of such numbers is like the introduction of complex numbers to factor polynomials into formal linear factors. He also compares them to the introduction of the common chord of two circles that do not intersect , and later, in a letter to Kronecker, introduces the most appealing analogy, comparing them to chemical elements. Since not all elements could be isolated at the time, they were indeed, as I understand, still regarded as ideal – as opposed to real – constructs.

To make the precise definition, he begins with the factorization of primes congruent to 1 modulo λ. Sometimes, and these were the calculations of the Königsberg paper, such primes admit a factorization

$$p = f(\alpha)f(\alpha^2)\dots f(\alpha^{\lambda-1}), \tag{E}$$

but not always. So he searches for a definition that will accomodate both an actual and a virtual factorization. If (E) is possible, then there is an integer u that satisfies the equation

$$u^{\lambda-1} + u^{\lambda-2} + \dots + u + 1,$$

and is such that $f(u) \equiv 0 \pmod{p}$. We have seen this, for we may take in our argument $\mathfrak{p} = (f(\alpha))$. More generally, any number in R is expressed as a polynomial $\Phi(\alpha)$ in α with integral coefficients and the prime $f(\alpha)$ divides $\Phi(\alpha)$if and only if $\Phi(u)$ is divisible by p.

He then observes that one can in fact introduce this as a definition. The complex integer $\Phi(\alpha)$ is divisible by the *prime divisor of p that belongs to $\alpha = u$* if $\Phi(u) \equiv 0 \pmod{p}$. At the same time, he observes that the definition is inadequate because it does not allow us to determine the power of the ideal factor dividing a given number, nor does it apply to the prime divisors of ordinary primes p that are not congruent to 1 modulo λ.

Continuing, he points out that no matter how we define prime factors of complex integers, each will have to be the divisor of some ordinary prime p. He lets p^f be the smallest power of p that is congruent to 1 modulo λ. Then he forms the periods

$\eta_1, \ldots, \eta_e$ and denotes the complex integer $c_1\eta_1 + \cdots + c_i\eta_i$ by $\Phi(eta)$. Then he states,

so giebt es unter den Primzahlen p, welche zum Exponenten f gehören, immer solche, die sich auf die Form

$$p = \Phi(\eta_1) \ldots \Phi(\eta_e)$$

bringen lassen, in welcher auch die e Factoren niemals eine weitere Zerlegung gestatten.

I confess that although I believe this statement when $f = 1$ because it follows from the generalizations of Dirichlet's theorem on primes in an arithmetic progression, I do not see any reason that it should be true in general, but that may be my ignorance.

In any case, when p can be so represented then $\Phi(\eta)$ is a prime factor of p as is the ideal $\mathfrak{p} = (\phi(\eta)$ in k. As we saw there are ordinary integers $u_1, \ldots, u_k$, such that each of the numbers $\eta_i - u_i$ is a multiple of $\Phi(\eta)$. Then Kummer writes something that at first made no sense to me,

Enthält nun irgend eine complexe Zahl $f(\alpha)$ den Primfactor $\Phi(\eta)$, so wird sie die Eigenschaft haben, für $\eta = u_k$, $\eta_1 = u_{k+1}$, $\eta_2 = u_{k+2}$, etc. congruent Null zu werden, für den Modul q. Diese Eigenschaft nun (welche eigentlich f besondere congruenzbedingungen in sich schließt deren Entwicklung zu weit führen würde) ist eine bleibende; auch für diejenigen Primzahlen q, welche eine Zerlegung in die e wirklichen complexen Primfactoren nicht gestatten.

It made no sense partly because I had not looked carefully enough at the last paper of the four that presumable includes the development that would be too much of a digression and partly because I had not looked carefully enough at the section on cyclotomy in the *Disquisitiones*. The third paper appeared in the same volume of Crelle's Journal as an introduction to the fourth and immediately before it. He does not warn the reader that he anticipates notation explained only in the longer paper. We can, in particular, ignore the subscripts.

Since I am more and more convinced that every mathematician should, at some point in his career, spend some time with the *Disquisitiones*, which has been translated into several languages, English among them, I recall the relevant fact from from Gauss and the use Kummer makes of it. It makes some of the earlier arguments more concrete. The pertinent theorem is in §347.

THEOREMA. *Si $F = \varphi(t, u, v, \ldots)$ est functio invariabilis algebraica rationalis integra f indeterminatarum t, u, v etc., atque substituendo pro his f radices in periodo (f, λ) contentas, valor ipsius F per praecepta art. 340 ad formam*

$$A + A'[1] + A''[2] + \text{etc.} = W$$

reducitur: radices quae in hac expressione ad eandem periodum quamconque f terminorum pertinent, coefficientes aequales habebunt.

I use Gauss's notation in the proof. He writes [p] and [q] for α^p and α^q, p and q being for a brief moment just two integers. If they belong to the same period of length e, then $q = pg^{\nu e}$, where g is now a generator of the nonzero residues modulo n (the notation Gauss uses for the prime λ, λ being for him any nonzero residue of n). He denotes the roots in the period (f, λ) by $[\lambda]$, $[\lambda'] = [\lambda g^e]$, $[\lambda''] = [\lambda g^{2e}]$ and so on. Now to calculate W, one substitutes α^λ, $\alpha^{\lambda'}$ and so on for $t, u, \dots$ multiplies the terms in each monomial together, in other words one adds the exponents, and according to the value of the resulting exponent modulo n, whether it is $0, 1, \dots, n-1$, one adds the coefficient of the monomial to the sum for A, A', A'', and so on. This is in particular true for the coefficient of $[p]$. On the other hand the sequence $\lambda g^{\nu e}$, $\lambda' g^{nue}$, $\lambda'' g^{nue}$ is modula n the sequence λ, λ', λ'' in a different order. Since φ is a symmetric function, the modified sequence can be used as well as the original to calculate W. It is clear that using the new sequence the constant term A in the expansion of W does not change, whereas the coefficient A' of [1] becomes the coefficient of $[g^{\nu e}]$, that of [2] becomes the coefficient of $[2g^{nue}]$ and so on. In particular, of if one prefers in general, the coefficient of $[p]$ becomes that of $[q]$. This is the theorem. Notice that if the symmetric polynomial has integral coefficients then the coefficients $A, A', A'', \dots$ will be integers.

Consider then the equation

$$(x - [\lambda])(x - [\lambda']) \cdots = x^f - ax^{f-1} + ax^{f-2} - + \cdots = 0$$

satisfied by the roots in a period. Because of the theorem, the coefficients are all expressible as linear combinations with integral coefficients of the e periods $\eta_1, \eta_2, \dots, \eta_e$.

In particular, for $[\lambda] = 1$, we have a relation that Kummer would write

$$\alpha^f + \mathbf{P}_1 \alpha^{f-1} + \mathbf{P}_2 \alpha^{f-2} + \cdots + \mathbf{P}_f = 0, \tag{F}$$

the coefficients being in $R \cap k$, thus linear combinations with integral coefficients of $\eta_1, \dots, \eta_e$.

As a consequence, any number

$$f(\alpha) = a + b\alpha + c\alpha^2 + \cdots + s\alpha^{\lambda-1}$$

in R can be written as

$$f(\alpha) = \varphi(\eta) + \alpha\varphi_1(\eta) + \alpha\varphi_2(\eta) + \cdots + \alpha^{f-1}(\eta), \tag{G}$$

where I ave used Kummer's notation for the coefficients, which are integral linear combinations of $\eta_1, \eta_2, \dots$.

We saw that a prime ideal $\mathfrak{p}$, thus in particular the ideal $(\Phi(\eta)$, defines a homomorphism $\eta_i \to u_i$ of $R \cap k$ into $\mathbb{Z}_p$. What Kummer, who proves the existence of this homomorphism in another way, means is that we substitute these values of

the η_i into the coefficients of $f(\alpha)$. Notice that substituting into (F) we obtain the equation of α modulo $\mathfrak{P}$, the extension of $\mathfrak{p}$ to K!

With the equation for α modulo $\mathfrak{P}$ at his disposal, Kummer can decide whether f is in $\mathfrak{P}$ simply by substituting for the coefficients of (G) their values modulo $\mathfrak{p}$, obtained by replacing eta_i by u_i. What he means by the word *bleibend* is that this definition remains valid even when the prime divisor is only *ideal*.

What these remarks do not give, as Kummer points out, is the power to which the ideal numbers introduced in this way divide a given number. So his final definition is different.

He considers forms $\psi(\eta)$, thus linear forms $a_1\eta_1 + a_2\eta_2 + \cdots + a_e\eta_e$. Thus for $\eta = \eta_1$,

$$\psi(\eta) = a_1\eta_1 + a_2\eta_2 + \cdots + a_e\eta_e,$$

while in general

$$\psi(\eta_i) = a_1\eta_{i+1} + a_2\eta_{i+2} + \cdots + a_e\eta_{i+e}.$$

If one likes, these are the conjugates of $\psi(\eta)$ under the Galois group of k over $\mathbb{Q}$, with which of course Kummer was familiar as a collection of concrete transformations but not as a concept. For each prime p with p of order f modulo λ, he chooses ψ such that

$$\mathrm{N}\,\psi(\eta) = \psi(\eta_1)\ldots\psi(\eta_e)$$

is divisible by p but not by p^2. Then he sets $\Psi(\eta) = \Psi(\eta_1) = \psi(\eta_2)\ldots\psi(\eta_e)$. He could define $\Psi(\eta_i)$ for any i in a similar way. Then the definition on which he bases his theorems is the following,

Wenn $f(\alpha)$ die Eigenschaft hat, daß das Product $f(\alpha).\Psi(\eta_i)$ durch p teilbar ist, so soll dies so ausgedrückt werden: es enthält $f(\alpha)$ den idealen Primfactor von p, welcher zu $u = \eta_i$ gehört. Ferner, wenn $f(\alpha)$ die Eigenschaft hat, daß $f(\alpha)(\Psi(\eta_i))^\mu$ durch p^μ theilbar ist, aber $f(\alpha)(\Psi(\eta_i))^{\mu+1}$ nicht theilbar durch $p^{|mu+1}$, so soll dies heißen: Es enthält $f(\alpha)$ den zu $u = \eta_i$ gehörigen idealen Primfactor von p genau μ mal.

Unfofrtunately, there is no more time to discuss Kummer's treatment of cyclotomic fields. What I want to observe now is that cyclotomic fields are class fields over $\mathbb{Q}$. So we have first to recall the earlier notions of a class field, due to Weber, Hilbert and Takagi. In connection with Weber, I observe that in the extensive bibliography to Hasse's famous *Klassenkörperbericht* in 1926/27 there are no papers of Weber given, although his name is mentioned. In a brief *History of class field theory* written much late, Hasse observes, however,

*The notion of **class field** is generally attributed to Hilbert. In truth this notion was already present in the mind of Kronecker and the term was coined by Weber, before Hilbert's fundamental papers appeared.*

I have not had a chance to look at Weber's papers to see what he had done.

Over an algebraic number field k, the notion of an ideal class is well known. Two integral ideals $\mathfrak{a}$ and $\mathfrak{b}$ belong to the same class is there are nonzero integers a and b such that $b\mathfrak{a} = a\mathfrak{b}$. It is a basic theorem that the ideal classes form a finite group. It can also be defined as the quotient of all fractional ideals, thus essentially ideals that can be represented as quotients $\mathfrak{a}/\mathfrak{b}$, by the group of principal fractional ideals (a). There is a more general notion.

Suppose first of all $\mathfrak{m}$ is an integral ideal. Every class has a representative in whose numerator and denominator only ideals relatively prime to $\mathfrak{m}$ appear. So we consider only those fractional ideals that are so represented. They form an infinite group. We divide by those principal ideals that are represented as a/b where both a and b are prime to $\mathfrak{m}$, a/b is positive in every imbedding $k \to \mathbb{R}$, and $a \equiv b \pmod{\mathfrak{m}}$. Once again, the result is a finite group, called I believe the *ray-class group*. It has the usual group of ideal classes as a quotient. The groups that figure in class-field theory are all quotients of the ray-class group. If $\mathfrak{m}$ divides $\mathfrak{m}'$ there is a homomorphism of the ray-class group modulo $\mathfrak{m}'$ onto the ray-class group modulo $\mathfrak{m}$. We shall call a group intermediate between the full group of ideals and the trivial ray-class modulo $\mathfrak{m}$, or equal to one of the extremes, an ideal-class group defined modulo $\mathfrak{m}$. Because of the homomorphisms just mentioned there is an equivalence relation on the collection of ideal-class groups and a given such group may have more than one conductor. Although that is not important for us here, it will always have a minimal conductor.

Consider, for example, $k = \mathbb{Q}$ and take $\mathfrak{m} = (\lambda)$, where λ is the prime we have taken from Kummer. Since (-1) is the trivial ideal, every ideal is represented by a positive number. Then the elements of the ray-class group modulo $\mathfrak{m}$ are represented by a/b with a and b prime to λ. Such a quotient represents the trivial class if and only if $a \equiv b \pmod{\lambda}$, thus if and only if $a/b \equiv 1 \pmod{\lambda}$. So the ray-class group is isomorphic to the multiplicative group of the nonzero residues modulo λ.

We have seen, except for the details of the proofs and a few exceptional primes that it was inconvenient to treat, that a prime ideal (p) in the field $k = \mathbb{Q}$ decomposes in the field $K = \mathbb{Q}(\alpha)$ into primes factors of degree 1 if and only if $p \equiv 1 \pmod{\lambda}$, thus if and only if the class of (p) in the ray-class group modulo $\mathfrak{m}$ is trivial. This makes K a class-field over k. The general definition given by Weber and used by Hilbert is simple enough to state. We take a conductor $\mathfrak{m}$ and an ideal-class group H of this conductor.

A relative Galois extension K of k is called a class-field for H if the prime ideals in k that are prime to $\mathfrak{m}$ and that decompose into K into prime ideals of degree 1 *are precisely the ideals in H.*

Thus the field $\mathbb{Q}(\alpha)$ is the class field for the ray-class group modulo λ.

Class-field theory was established in its first form by Takagi. There are several basic theorems, among them the following.

1) There is a unique class field K associated to each ideal-class group H.

2) The class field K is abelian over k and the Galois group of K over k is isomorphic to the quotient of the group of all ideal classes prime to $\mathfrak{m}$ by H.

3) If f is the order of $\mathfrak{p}$ modulo H, thus $\mathfrak{p}^f$ the smallest power of $\mathfrak{p}$ contained in H, then $\mathfrak{p}$ decomposes in K into prime factors of relative degree f.

4) Every abelian extension of k is a class field to some ideal-class group.

A somewhat strange feature of the theory as it was created by Takagi is that the isomorphism of (2) is not explicit.

The isomorphism was made explicit by Emil Artin, whose contributions led to a recasting of the theory, largely, I believe, by Chevalley. The ideas of Chevalley, preceded by those of Herbrand and followed much later by ideas of Shafarevitch and Weil, have led to a theory with quite a difference emphasis and with quite different proofs in which idèles and group cohomology figure prominently. For better or worse they largely obscure the original definition of Weber as well as the original immediacy of the constructions.

There is first of all a major change in the proof, which results, both in its new form and in the old, from a complicated chain of arguments. As an essential part of these arguments, there are two fundamental inequalities, the first and the second. The curious thing is that the first in the Takagi argument becomes the second in the modern argument, and the first in the modern argument is the second in the Takagi argument. Another important difference is that one of the two, the first, is analytic in the Takagi argument, whereas one of the two, still the first, is cohomological in the modern argument. The other, the second in both forms, is based on a careful study of a specific collection of extensions, the Kummer extensions.

A central concern of the general theory of automorphic forms, of which class field theory can now be regarded as a particular case are general reciprocity laws that may be regarded as pretty much the ultimate forms of those discovered by Artin. There is little doubt that the trace formula will, in some form or other be an essential element of any arguments used to establish them. I recall that the trace formula, apart from the very serious analytic difficulties attached to it, expresses a simple formal principle, the same formal principle as the Frobenius reciprocity law. Since the theory of automorphic forms is the study of the action of a group G on functions on $\Gamma\backslash G$, it is, therefore, hardly surprising that it will be used in any serious study of them. There are, however, at least two quite different ways in which it might be used. The first has been with us for two or three decades, has had some important successes, and is by no means exhausted. Although the first way can seldom be carried out effectively without a great deal of analytic labor, the basic principle that it exploits is algebraic.

To explain it, I recall the proof of the trace formula for the special case that the quotient $\Gamma\backslash G$ is compact. Then there are no serious analytic difficulties to overcome

when establishing it. The operator $R_f\phi(g) = \phi'(g)$, with $\phi'(g) = \int_G \phi(gh)f(h)dh$ on $L^2(\Gamma\backslash G)$ is of trace class if f is smooth with compact support. Since

$$\int_G \phi(gh)f(h) = \int_{\Gamma\backslash G} \phi(h) \sum_{\Gamma} f(g^{-1}\gamma h)dh,$$

the kernel is clearly $\sum f(g^{-1}\gamma h)$. To obtain the trace of R_f, we integrate the kernel over the diagonal, obtaining a sum over conjugacy classes in Γ,

$$\sum_{\{\gamma\}} \int_{\Gamma_\gamma\backslash G} f(g^{-1}\gamma g)dg = \sum_{\{\gamma\}} \mu(\Gamma_\gamma\backslash G_\gamma) \int_{G_\gamma\backslash G} f(g^{-1}\gamma g)dg.$$

The symbol μ is the Haar measure on the pertinent quotient and the subscript γ denotes the centralizer of γ. The formula is the equality of this sum with the sum of the eigenvalues of R_f.

This is in essence what the trace formula yields when the quotient $\Gamma\backslash G$ is not compact. A typical application of the trace formula of the first type is to a comparison of the spectrum, thus of the eigenvalues of R_f and $R_{f'}$, where there are two different groups G and G', with, therefore, different discrete subgroups Γ and Γ', but where for some largely algebraic reason there is a close relation between the conjugacy classes in Γ and in Γ' and, simultaneously, between those in G and G'. Then f and f' are so chosen that the orbital integral of f at γ is equal to the orbital integral of f' at γ'. Then

$$\sum_{\{\gamma\}} \mu(\Gamma_\gamma\backslash G_\gamma) \int_{G_\gamma\backslash G} f(g^{-1}\gamma g)dg$$

can be compared with

$$\sum_{\{\gamma'\}} \mu(\Gamma'_{\gamma'}\backslash G'_{\gamma'}) \int_{G'_{\gamma'}\backslash G'} f(g^{-1}\gamma' g)dg.$$

The second way in which it might be used has only recently been proposed. It has been exploited to prove some known results, but it has not yet been used in any way to obtain new results. If it works, it will be much more powerful than the first way. Indeed the first way or method would be an element of the second because the second envisages an extra step, namely the introduction of an additional limiting process for both G and G'. The second method envisages incorporating the methods of analytic number theory in the trace formula not simply to a much greater extent than before but, in fact, for the first time. At the moment, we are not beyond the stage of numerical experimentation. Although the experiments are still, I have to admit, inconclusive, I hope to describe briefly some promising signs at the very end of these lectures.

What I want to do now, as a transition to automorphic forms on reductive groups is, first, to describe two features of the proofs of class-field theory before Artin and Chevalley, the analytic aspect of the proof of the first inequality, and the explicit counting, which is a feature in one way or another of both the old and the new class-field theory. In both it appears in the proof of the second inequality, but, I recall, these are not the same inequalities in the two theories. Then I want to recall the reciprocity law of Artin, which had an enormous effect on the formulation of the laws that we are trying to establish in the general theory of automorphic forms – with, I stress, some success.

Takagi's definition of a class field is different from that of Weber and Hilbert, although the concept is ultimately the same. Suppose K is an extension field of k, finite over k. If $\mathfrak{m}$ is any integral ideal of k let $H_{\mathfrak{m}}$ be the collection of all ray-classes modulo $\mathfrak{m}$ that contain ideals that are norms of ideals in K. Let $h_{\mathfrak{m}}$ be its index in the full group of all ray-classes. Then the first inequality states

If K is Galois over k and of degree n then, for every integral ideal $\mathfrak{m}$

$$h_{\mathfrak{m}} \leq n.$$

This can be proved with the help of basic analytic properties of the L-functions associated to characters χ of the ray-class group modulo $\mathfrak{m}$. These are the functions

$$\begin{aligned} L(s,\chi) &= \sum_{(\mathfrak{a},\mathfrak{m})} \frac{\chi(\mathfrak{a})}{\mathrm{N}\mathfrak{a}^s} \\ &= \sum_{\mathfrak{h}\in H_{\mathfrak{m}}} \chi(\mathfrak{h}) \sum_{\mathfrak{a}\in\mathfrak{h}} \frac{1}{\mathrm{N}\mathfrak{a}}, \end{aligned}$$

$\mathfrak{h}$ denoting a ray-class modulo $\mathfrak{m}$.

By what is in essence a simple geometric argument based on Dirichlet's unit theorem, one shows first of all that these series converge for $s > 1$, secondly that

$$\lim_{s\to 1}(s-1)\sum_{\mathfrak{a}\in\mathfrak{h}} \frac{1}{\mathrm{N}\mathfrak{a}} = c,$$

where c is a constant that depends on k and $\mathfrak{m}$.

The ray-class group can be replaced by any ideal-class group H defined bddmodulo some $\mathfrak{m}$. Then h will be a coset modulo H and the group of all idealsthe constant c will be replaced by a constant that depends on H alone. As a result, if $\chi = \chi_0$ is the trivial character on cosets modulo H, then

$$\lim_{s\to 1}(s-1)L(s,\chi) = c\cdot h,$$

where h is the number of classes modulo H. If χ is not trivial then

$$\lim_{s\to 1} L(s,\chi)$$

is finite, although it may be 0.

The function $L(s,\chi)$ is given as an Euler product,

$$L(s,\chi) = \prod_{(\mathfrak{p},\mathfrak{m})=1} \frac{1}{1-\frac{\chi(\mathfrak{p})}{\mathrm{N}\mathfrak{p}}},$$

so that

$$\log L(s,\chi) = \sum_{\substack{(\mathfrak{p},\mathfrak{m})=1\\ m=1,2,\dots}} \frac{\chi(\mathfrak{p}^m)}{m\,\mathrm{N}\mathfrak{p}^{ms}} = \sum_{(\mathfrak{p},\mathfrak{m})=1} \frac{\chi(\mathfrak{p})}{\mathrm{N}\mathfrak{p}^s} + g(s,\chi), \tag{H}$$

where $g(s,\chi)$ is analytic for $\Re s > \frac{1}{2}$. The sum over higher powers of primes is finite for $s > 1/2$ for the usual reasons.

If we sum (H) over the characters χ modulo H and divide by h, we obtain

$$\sum_{\substack{(\mathfrak{p},\mathfrak{m})=1\\ \mathfrak{p}\in H}} \frac{1}{\mathrm{N}\,\mathfrak{p}^s} = \frac{1}{h}\log\frac{1}{s-1} + f(s), \tag{I}$$

where

$$f(s) = \frac{1}{h}\log\{(s-1)\prod_{\chi} L(s,\chi)\} = \frac{1}{h}\log\{(s-1)L(s,\chi_0)\}\prod_{\chi\neq\chi_0}\log L(s,\chi), \tag{J}$$

a function that approaches a finite value $f(1)$ if none of the $L(s,\chi)$, $\chi\neq\chi_0$, are 0 and $-\infty$ if some are.

We apply these considerations first to k and to the ideal-class group $H_{\mathfrak{m}}$ in k that we attached to K, thus the group of all ray-classes that contain norms of ideals in K. The left side is then

$$\sum_{\mathfrak{p}\in H_{\mathfrak{m}}} \frac{1}{\mathrm{N}\mathfrak{p}^s}. \tag{K}$$

For an arbitrary field k, thus in particular, the extension field K, we can apply the discussion to the ideal-class group of all ideal classes, defined modulo (1). Then the final product in (J) is empty and the function $f_1(s)$ that appears on the left has a positive limit as $s\to 1$. Thus

$$\sum_{\mathfrak{P}_1} \frac{1}{\mathrm{N}\mathfrak{P}_1^s} = \log\frac{1}{s-1} + f_1(s). \tag{L}$$

We apply this to K. then we can discard on the left those ideals whose relative degree is not 1 because as usual we have a sum that is dominated by a constant times

$$\sum_p \frac{1}{p^{2s}}.$$

Moreover, as $[K : k] = n$, each ideal $\mathfrak{p}_1$ in k, at least each ideal that does not divide the discriminant and that factors into prime ideals of degree 1 over is divisible by exactly n prime ideals $\mathfrak{P}_1$. Consequently, retaining only the ideals of degree 1 over k in (L) and disregarding the effect of a finite number, for that can always be incorporated into $f_1(s)$, we have

$$\text{(M)} \qquad \sum_{\mathfrak{p}_1} \frac{1}{\mathrm{N}\mathfrak{p}_1^s} = \frac{1}{n} \log \frac{1}{s-1} + g_1(s),$$

where $g_1(s)$ approaches a finite value as $s \to 1$.

Every ideal that appears in the sum of M also appears in (K), so that we obtain a nonnegative quantity when subtracting from (M) from (K). Consequently

$$0 \leq \big(\frac{1}{h_{\mathfrak{m}}} - \frac{1}{n}\big) \log \frac{1}{s-1} + f(s) - g_1(s).$$

Finally $f(s) - g_1(s)$ either remains finite or approaches $-\infty$ as $s \to 1$. Since $\log(1/(s-1))$ approaches $+\infty$, we conclude that

$$\frac{1}{h_{\mathfrak{m}}} - \frac{1}{n} \geq 0.$$

This is the asserted inequality.

What I want to stress is that we have used in this argument, the logarithms of L-functions. We could also use the derivatives, or the negative derivatives of these functions. The argument would be essentially the same, although we might need more information about $f(s)$ near $s = 1$, for example, that $f(s)$ is differentiable at $s = 1$. What I will want to suggest later is that the use of logarithms of L-functions or, more conveniently, the logarithmic derivatives of L-functions may be an important tool in the investigation of various general conjectures in the modern theory of automorphic forms. I have recalled this proof to convince you that this would not be entirely novel.

At the same time, the arguments of Takagi, his predecessors, and his successors require some explicit information about specific abelian extensions. This explicit information is the number-theoretical ingredient in the arguments. At the moment, it is not at all clear what its analogue will be in the general theory.

Recall that according to the definition of Takagi, a relative Galois extension K/k of degree n is a class-field associated to the ideal-class group H defined modulo $\mathfrak{m}$ if H and the ideal-class group $H_{\mathfrak{m}}$ are equal and if, in addition, $h_{\mathfrak{m}} = h = n$.

In the construction of the theory, there are various reductions, and the construction is finally reduced to showing that for every prime number l and every field k that contains the lth roots of unity, every relatively cyclic extension of degree l is a class field and that to every ideal-class group of index l there is associated a class-field. Since, as one shows, this correspondence is unique, it is pretty much a matter of counting, provided one establishes beforehand that every abelian extension is the class-field associated to some H. Finiteness of the objects counted is assured by fixing discriminants of the fields and conductors of the ideal-class groups (the smallest ideal $\mathfrak{m}$), or at least bounds on them.

The abelian extensions of degree l of k are given explicitly as $k(\sqrt[l]{\alpha})$, where $\alpha \in k$. Moreover $k(\sqrt[l]{\alpha}) = k(\sqrt[l]{\beta})$ if and only $\alpha^i/\beta^j \in k^l$, with i and j prime to l. To show that they are class fields, it has to be shown that $h_{\mathfrak{m}} = l$ for a suitable $\mathfrak{m}$. In view of the inequality $h_{\mathfrak{m}} \leq l$ already proved, it is enough to show that $h_{\mathfrak{m}} \geq l$, for a suitable $\mathfrak{m}$. This is, in the original class-field theory, the second inequality and is a serious cohomological calculation. This inequality proven, we are left with counting on the one hand, the order not of k/k^l, which is of course infinite, but of the subgroup obtained from elements that are units outside a given finite set of primes and the number of ideal-class groups of index l whose conductors are divisible by a given set of primes. This counting argument contains some subtleties that we can ignore here.

As I said, the general ideas are still inchoate. In particular, I have not yet seen how and where a counting argument might occur. The first step is perhaps to see more clearly where analytic arguments might take us. If I had had more time, my impulse would have been to reflect on the one place in which counting arguments have appeared up to the present in the theory of automorphic forms on general groups, namely in Wiles's proof of the Shimura-Taniyama-Weil conjecture.

The construction of class-field theory took a different form in the basic paper of Chevalley *La théorie du corps de classe (1940)*, but I want to stress here an entirely different consequence of the reformulation of the theory by Artin and Chevalley.

Whereas in the early theory, the reciprocity laws took to some extent a secondary place, in the theory following Artin, the reciprocity law between the Galois group $\mathrm{Gal}(K/k)$ and the classes modulo the associated ideal-class group H, was introduced into the very foundations of the theory. So was its consequence, that every abelian Artin L-function was equal to one of the L-functions attached by Hecke to what he called a *Grössencharakter*. This was, and remains, the only way to show that abelian Artin L-functions can have an analytic continuation.

This reciprocity law and its analytic consequence were to a substantial extent the inspiration for functoriality, but only after they were combined with another ingredient, the use of idèles by Chevalley. This is, to some degree, just another way of introducing ideal-class groups, but in such a way that there is no need to employ an equivalence relation to identify groups defined with respect to different $\mathfrak{m}$.

Before I pass to the functoriality and general reciprocity laws, let me recall that the problem of analytically continuing the Artin L-functions to the entire plane with

only the poles on the real line determined by the Γ-factors was first raised by Artin's paper of 1923. Although it was solved for abelian Artin L-functions a few years later when Artin established his reciprocity laws and although Richard Brauer's theorem on group characters, proved in 1946, establishes its analytic continuation as a meromorphic function, the problem remains open, in spite of some progress, until this day. Our view, at least my view, of the problem changed in 1967. To give you some idea of the notions prevailing at that period, I quote a statement made by Artin at the Princeton University Becentennial Conference on Mathematics in 1946.

In the minutes of the discussions we read,

... Brauer's result represents a decisive step in the generalization of class-field theory to the non-Abelian case, which is commonly regarded as one of the most difficult and important problems in modern algebra.
Artin stated that, "My own belief is that we know it already, though no one will believe me – that whatever can be said about non-Abelian class field theory follows from what we know now, since it depends on the behaviour of the broad field over the intermediate fields – and there are sufficiently many Abelian cases." The critical thing is learning how to pass from a prime in an intermediate field to a prime in the large field. "Our difficulty is not in the proofs, but in learning what to prove."

Fortunately, it has not turned out to be so simple.

Adèles - influence of Chevalley, Weil, Tamagawa - also Artin - Tate - also Herbrand

- algebraic groups

- their structure (Galois cohomology)

- Siegel's development of the reduction theory of Gauss, Hermite, Minkowski and of formulas representation of integers by quadratic forms, formulas due to Eisenstein, H. J. L. Smith, Minkowski, led to the use of general (reductive) algebraic groups in the theory of automorphic forms

- Hecke's introduction of L-functions into the theory of automorphic forms

- representation theory (addition from theoretical physics) - Dirac, Wigner $\rightarrow$ Harish-Chandra

- spectral theory (Maass, Selberg)

Basic objects

$G(\mathbb{A})$ G : reductive algebraic group, typically $\mathrm{GL}(n)$

$\mathbb{A}$: ring of adèles

$a = a_\infty, a_2, a_3, \ldots$

$a_\infty \in G(\mathbb{R}), \quad a_2 \in G(\mathbb{Q}_2), \quad a_3 \in G(\mathbb{Q}_3), \ldots$

$G(\mathbb{Q})$ quotient $G(\mathbb{Q})\backslash G(\mathbb{A})$

φ : function on $G(\mathbb{Q})\backslash G(\mathbb{A})$

$G(\mathbb{A})$ acts on these

π : irreducible representation

$= \otimes\, \pi_v = \pi_\infty \otimes \pi_2 \otimes \pi_3 \ldots$

π_p : unramified for a.a. p

$\pi_p \leftrightarrow$ conjugacy class $\{A(\pi_p)\}$ in dual group

Dual group: ${}^L G : \widehat{G} \rtimes \mathrm{Gal}(K/F)$

F : ground field : $\mathbb{Q}$ or a finite algebraic number field

$\widehat{G}$: connected complex group

$G = \mathrm{GL}(n)$ $\qquad\qquad$ $\widehat{G} = \mathrm{GL}(n, \mathbb{C})$

Note: $\quad G = \{1\} \Rightarrow \widehat{G} = \{1\}$

${}^L G = \ \mathrm{Gal}(K/F)$

General automorphic L-function

$$L(s, \pi, \rho) = \prod_{p \notin S} L(s, \pi_p, \rho)$$

ρ : finite-dimensional complex representation of ${}^L G$.

$$L(s, \pi_p, \rho) = \frac{1}{\det\left(1 - \frac{\rho(A(\pi_p))}{p^s}\right)}$$

Number field: $p \to N_{\mathfrak{p}}$

Basic analytic properties of these functions can be treated when $G = \mathrm{GL}(n)$ and ρ is the standard representation.

Generalization of Artin reciprocity law. = *Functoriality in the L-groups*

Two groups G and G'

$$\begin{array}{ccc} \phi : \ {}^L G & \longrightarrow & {}^L G' \\ & \searrow \quad \swarrow & \\ & \mathrm{Gal}(K/F) & \end{array}$$

π : automorphic representation of G

: $\{A(\pi_p)\}$

$\{A'_p\} = \{\phi(A_p)\}$.

Assertion: $\exists\ \pi'$ automorphic representation of G' such that $A'_p = A(\pi'_p)$ for almost all p. For Artin reciprocity take

$$G = \{1\} \qquad\qquad G' = \ \mathrm{GL}(1)$$

Two touchstone cases

a) $G = \{1\}$ $\qquad$ $G' = \mathrm{GL}(n)$

Would yield analytic continuation of Artin L-functions.

b) $G = \mathrm{GL}(2)$ $\qquad$ $G' = \mathrm{GL}(n)$

$^L G = \mathrm{GL}(2, \mathbb{C})$

ϕ : irreducible representation of GL(2) of degree n.

Yields various forms of Ramanujan conjecture.

Note - there are spectral - theoretic questions

$$G \to \pi \to A(\pi_p) - \quad \text{algebraically}$$

convenient way of expressing spectral-theoretic properties of functions in space of π.

$G' \to \pi'$? - does there exist π' with spectral-theoretic properties given by $\phi(A(\pi_p)) = A'_p$?

.

.

Trace formula (Selberg-Arthur)

f : function on $G(\mathbb{A})$ $\qquad$ $f(g) = \prod_v f_v(g_v)$

φ : function on $G(\mathbb{Q})\backslash G(\mathbb{A})$

$\varphi \to R(f)\varphi(g) = \int_{G(\mathbb{A})} \varphi(gh) f(h) dh$

$$\boxed{\text{trace } R(f) = \sum_{\pi} \text{ trace } \pi(f)}$$

An approximate statement. Trace formula expresses this as a sum over *conjugacy classes* in $G(\mathbb{Q})$.

Thus the analytic information is expressed by arithmetic information.

First method for using the trace formula - choose f and f', f in $G(\mathbb{A})$, f' in $G'(\mathbb{A})$

$$\text{trace } R(f) = \sum_{\{\gamma\}} \mu(G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A})) \int f(g^{-1}\gamma g)dg$$

$$\text{trace } R(f') = \sum_{\{\gamma'\}} \mu(G'_\gamma(\mathbb{Q})\backslash G'_\gamma(\mathbb{A})) \int f'(g^{-1}\gamma g')dg$$

used for base change for cyclic extensions, used by Arthur to compare automorphic forms in classical groups with automorphic forms on $\mathrm{GL}(n)$.

Needs - relation between conjugacy classes γ and conjugacy classes γ' (uses also *twisted* trace formulas)

- relation between orbital integrals.

Note - These relations are deduced from Galois cohomology, thus in essence from the abelian theory. So they conform to Artin's prediction. They are essentially(!) algebraic.

Yield important results. For example, *enough for Fermat's theorem.* Nevertheless in comparison to what is expected they yield *very limited* results

.

.

Can we combine analytic and algebraic methods?

Logarithm:

$$\ln \quad L(s, \pi, \rho).$$

Structure of a possible argument

Logarithmic derivative

$$-\frac{L'}{L}(s, \pi, \rho).$$

Discard π that are *not of Ramanujan type*

Expect: $L(s, \pi, \rho)$ zero-free, pole-free for Re $s > 1$, perhaps poles (finite in number) on Re $s = 1$.

Let $m_\pi(\rho)$ be the order of the pole of $s = 1$. Then $m_\pi(\rho) \geq 0$. (This is of course not proved, since we do not even have the analytic continuation in general. We are trying to

find an approach to the problems.)

$$m_\pi(\rho) \text{ is residue of } -\frac{L'}{L}(x,\pi,\rho) \text{ at } s=1.$$

$$m_\pi(\rho) = \lim_{X\to\infty} \frac{\Sigma_{p<X} \ln p \ \mathrm{tr}(\rho(A(\pi_p)))}{X} \tag{A}$$

Might want to take a weighted average!

Expectation: Given π there is a subgroup

$$\begin{array}{ccc} {}^{\lambda}H_\pi & \subseteq & {}^{L}G \\ & \searrow \ \swarrow & \\ & \mathrm{Gal}(K/F) & \end{array}$$

such that $m_\pi(\rho)$ is the multiplicity with which the trivial representation is contained in the restriction of ρ to λ_{H_π}.

This would be a strengthened form of functionality.

First question - Is the definition (A) a way of getting our hands on something that could be called $m_\pi(\rho)$.

Can find a formula for

$$\sum_\pi \frac{\sum_{p<X} \ \mathrm{tr}(\rho(A(\pi_p)))}{X} \tag{B}$$

The outer sum is over automorphic representations with multiplicities, with representations not of Ramanujan type excluded.

BETTER S finite set of places, including infinite places.

$$\sum_\pi \left\{ \frac{\Sigma_{p<X} \ln p \ \mathrm{tr}(\rho(A(\pi_p)))}{X} \right\} \prod_{v\in S} \mathrm{tr}\left(\pi_v(f_v)\right) \tag{C}$$

On the one hand

$$\to \sum_{\pi} m_{\pi}(\rho) \prod_{v \in S} \operatorname{tr}(\pi_v(f_v)) = \sum_{\pi} \frac{1}{X} \sum_{p<X} \ln p \ \operatorname{tr}(\pi(f^{\rho,p})) \tag{D}$$

$$f^{\rho,p}(g) = \prod_{v} f_v(g_v)$$

$v \in S \qquad f_v$ given f_v

$v = p \qquad f_v = f_p^{\rho}$

$$\operatorname{tr}\ \pi_p(f_p^{\rho}) = \begin{cases} 0 & \pi_p \ \text{not unramified} \\ \operatorname{tr}(\rho(A(\pi_p))) & \pi_p \ \text{unramified} \end{cases}$$

$$v \notin S \neq p \qquad \operatorname{tr}\ \pi_v(f_v) = \begin{cases} 0 & \pi_v \ \text{ramified} \\ 1 & \pi_v \ \text{unramified.} \end{cases}$$

Thus the sum in (C) is over π unramified outside of S.

So question becomes:

Does

$$\frac{1}{X} \sum_{p<X} \ln p \sum_{\pi} \operatorname{tr}\ \pi(f^{\rho,p})$$

have a limit and can this limit be understood well enough to make comparisons possible? Recall that limit is to give

$$\sum_{\pi} m_{\pi}(\rho) \prod_{v \in S} \operatorname{tr}(\pi_v(f_v))$$

AT THIS POINT WE HAVE REACHED A PROBLEM WE CAN INVESTIGATE.

Take $G = \mathrm{GL}(2)$, $F = \mathbb{Q}$, and $S = \{\infty\}$

The only variable element is f_∞. Take it to be positively homogeneous. Invariant distributions evaluated at f_∞ are functions of two functions $\psi_\pm$ on two lines

$$\underline{\qquad\qquad} \ | \ \underline{\qquad\qquad} \ | \ \underline{\qquad\qquad} \ +$$

$$-1 \qquad\qquad 1$$

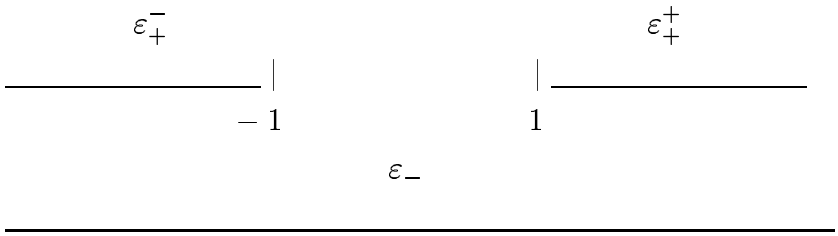
$$\underline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad} \ -$$

$$x = \frac{\text{trace} A}{2\sqrt{|\det A|}} \qquad \pm = \quad \text{sgn det } A$$

$$\psi(x) = \int f_\infty(g^{-1}\gamma g)dg$$

$$\gamma = A.$$

Measures on $\pm$ lines yield distributions on $\psi_\pm$ and thus invariant distributions. For example, Lebesgue measures



$$\varepsilon_+^- = \varepsilon_+^+ = \varepsilon_- = 4$$
$$\varepsilon_+^- = \varepsilon_+^+ = -\varepsilon_- = 4$$
$$\varepsilon_+^- = \varepsilon_+^+ = 4 \quad \varepsilon_- = 0.$$

These are characters of three irreducible representatins $\pi(\sigma)$ of GL(2) corresponding to three representations of Galois group $\text{Gal}(\mathbb{C}/\mathbb{R})$

$$\sigma : \varepsilon \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \varepsilon \to \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \varepsilon \to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note if ρ is not trivial then $m_\pi(\rho)$ is usually 0. Expected that it is not 0 only if

a) π of dihedral type

b) $\pi = \pi(\sigma)$, σ representation of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$

So we can expect to see measures that are linear combinations of the three measures just described appearing frequently.

.

.

TRACE FORMULA - Take $m = 1$ (no real information to be gained but lots of experience)

Parabolic terms - expressible in terms of the three simple measaures.

$$\text{Trace} = \sum_{N=\pm 4p^m} \sum_r \mu_D \frac{\psi_\pm\left(\frac{r}{\sqrt{|N|}}\right)}{p^{m/2}} \sum_{f|s} f \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right)$$

r : integer, m fixed, for a given p

$$r^2 - N = s^2 D$$

D : fundamental discriminant, $D \equiv 0, 1 \pmod 4$

s integer, > 0, as large as possible

s and D are functions of r and N

μ_D : invariant of field $\mathbb{Q}(\sqrt{D})$, class number, regulator

Two modifications to be made to formula

1) remove contributions of 1-dim, representations - they are not of Ramanujan type.

2) Replace μ_D by its analytic representation

$$\mu_D = \sum_{n=1}^{\infty} \left(\frac{D}{n}\right) \varphi(D, n)$$

$\boxed{\varphi(x, n)}$

$$x < 0 \qquad = \pi \ \text{Erfc}\left(\frac{n\sqrt{\pi}}{\sqrt{|x|}}\right) + \frac{\sqrt{|x|}}{n} \exp\left(\frac{-\pi_n^2}{x}\right)$$

$$x > 0 \qquad = \frac{\sqrt{x}}{n} \ \text{Erfc}\left(\frac{n\sqrt{\pi}}{\sqrt{|x|}}\right) + E_1\left(\frac{\pi n^2}{x}\right)$$

$$E_1(y) = -\gamma - \ln y + \sum_{k \geq 1} (-1)^{k-1} \frac{y^k}{k! k}$$

$$\sum_{\substack{f,n\\(f,n)=1}} 2\left\{\sum_{r,f',\pm} f\left(\frac{D}{nf'}\right)\varphi(D,nf')\frac{\psi_\pm(x_r)}{\sqrt{|N|}}\Phi\right.$$

$$\left.-\sqrt{|N|}\sum_{\pm}\varepsilon_{n,f}(N)\int\psi_\pm(x)\sqrt{|x^2\mp 1|}dx\right\}$$

$$x_r=\frac{r}{2\sqrt{|N|}}$$

$$\Phi=\prod_{q|f}\left(1-\frac{\left(\frac{D}{q}\right)}{q}\right)$$

Need to consider the average over $p<X$ with weight $\ln p$.

$$\frac{\Sigma_{p<X}\ln p\ \Sigma_{f,n}\xi^m_{n,f}(p)}{X}=\frac{\Sigma_{p<X}\ln p\ \Xi^m_{(p)}}{X}=\Theta^m(X)$$

$$\Xi^m(p)=\Sigma_{f,n}\xi^m_{n,f}(p).$$

Set

$$\theta^m_{n,f}(X)=\frac{\Sigma_{p<X}\ln p\ \xi^m_{n,f}(p)}{X}$$

$$\Theta^m(X)=\sum_{n,f}\theta^m_{n,f}(X)$$

Study asymptotic behavior of $\theta^m_{n,f}(X)$.

Three ranges.

$$n\ll\sqrt{X},\qquad n\sim\sqrt{X},\qquad n\gg\sqrt{X}$$

1) $n\ll\sqrt{X}$ both terms very small

2) $n\sim\sqrt{X}$ both terms $O(1)$

3) $n\gg\sqrt{X}$ both terms $\sim\sqrt{|N|}$

Examine separately - so far I have only examined (3). Thus asymptotic behavior of $\theta_{n,f}(X)$ for large X. *Only numerically.* Take $f=1$. Generally took $m=1$ because numerics are simpler and faster.

$$\xi^m_{n,f}(N)\qquad N=\pm 4p^m$$

Highly irregular form of p (m fixed). Is it

$$\xi^m_{n,f}(N) : O(\ln^2 p) \quad ???$$

LAPID-SARNAK: $O(p^{1/3})$

There are reasons to be sceptical about anything better than $O(p^{1/4})$ but experiments suggest nevertheless $O(\ln^2 p)$.

Source of irregularity

$$r^2 - N \to \frac{r^2 - N}{s^2} = D.$$

One has to sieve to find S

Take average.

$$\theta^m_{n,f}(X) = \alpha + \rho \ln X + \gamma \ln^2 X$$

$\alpha,\ \beta,\ \gamma$: measures.

Calculate measures of intervals of length 1 and length .1